

THIS WEEK

EDITORIALS

WORLD VIEW Science on the silver screen remains simplistic **p.139**

CLIMATE Global dimming offsets Arctic warming, for now **p.140**



LANGUAGE Chimps alter grunts to talk to new neighbours **p.141**

Telling stories

The UK Research Excellence Framework's focus on impact is a useful reminder of all the ways that science can help society — both economically and by other means.

People embark on a career in science for many reasons. Some want to improve the world, others to understand how it works. But how many foresee that their work will help to resurrect a sixteenth-century English warship?

In the language of twenty-first-century science, such research has a new label: impact. Hundreds of thousands of people, after all, have queued to see the Tudor timbers of the partially restored *Mary Rose*, salvaged from the sea floor, and now on display in a museum in Portsmouth, UK.

They do so thanks to the efforts of physicists, who tested radar imaging on the wreck site; marine biologists, who spotted borer worms still living in the timber; and chemists, who created nanoparticles to prevent the waterlogged wood being damaged by bacterial action. Once artefacts had been brought up from the wreck, materials scientists examined the corrosion on Tudor cannon balls; biomechanics experts analysed the arm bones of Tudor archers; and archaeologists inspected skulls to reconstruct the faces of the *Mary Rose's* crew. And all of this work was paid for — at least partially — by the British taxpayer, as part of UK investment in publicly funded science.

If scientists were once coy about the good work that they do, they cannot now afford to be. In fact, the British system now demands that they boast of the impact their research has on society. For the first time, the mammoth multi-year assessment of UK university research, used to help rank institutions and allocate grants, included judgements of such impact. This is a good thing.

The case studies and reports from this Research Excellence Framework assessment have now been published, providing a compendium of some 7,000 stories of good done, lives saved and ancient warships fixed up. As we discuss on page 150, scholars of research impact are rubbing their hands together at the thought of analysing the stories. Preliminary text-mining suggests that across many disciplines, studies strewn with words justifying the significance or reach of the work — such as 'million', 'major' and 'global' — tended to score more highly than narratives that over-used words such as 'research', 'university' and 'impact'.

Conventional measurements of research impact beyond academia seek hard data, not stories. They typically revolve around econometric models that try to capture the financial return of investing in science, or count small slices of quantifiable business activity, such as patents or spin-out companies. To be sure, there are plenty of those examples in the case studies. But taken as a whole, the narratives remind us of the many broader ways in which taxpayer-funded research 'pays back' on its investment — and that hard metrics are not the only way to capture this.

Indeed, one benefit of the focus on broad impact is that individuals and institutions that do good work that makes a positive difference to people's lives, society and the economy earn recognition — and motivation — even if they are not producing profound scientific insights.

There are some practical difficulties of running such an assessment, especially for the first time. Some researchers say that although they

are pleased to see the results, it was not worth the burden on academics' time and university budgets involved in collecting the case studies.

And it is true that, although a large set of good-news stories makes a valuable collection to dip into for advocacy purposes, the narratives from this particular exercise do not give a comprehensive view. Universities had to submit only a few of their best examples (and according to the data, many may have minimized the number of staff members

"If scientists were once coy about the good work they do, they cannot now afford to be."

whose work was submitted, so as to cut down on the number of case studies that they had to provide). Another problematic area concerns the difficulty of grading case studies when many different universities might each claim an influence on a final product (for example, a drug brought from bench to bedside).

These are teething troubles. The decision of the UK funders to grade the case studies, and to use the scores to help them to decide the destination of £2 billion (US\$3 billion) in performance-linked annual funding, meant that universities across the country have taken the exercise seriously. The result is a reminder of the many ways in which publicly funded research benefits society in the United Kingdom and beyond.

It demonstrates one other important point. Although the 'impact agenda' may focus minds and give universities and funders another way to make science tangible and measurable, the UK exercise shows that academics had been committing to impact long before it became a buzzword. The impact claimed is recent, within the past 5 years or so, but the research on which that impact is based is often up to 20 years old.

The focus on impact is a new thing, in other words — but the creation of impact is not. The more visible those impacts become, the better for all concerned. ■

Spot the difference

The US measles outbreak highlights why most states should reconsider their vaccination rules.

Over the past decade, increasing numbers of US parents have chosen not to vaccinate their children against diseases such as whooping cough, mumps and measles. The consequence has been a periodic return of these historical scourges, in localized outbreaks of a few dozen to a few hundred people. These episodes often appear in local news reports, some of which warn that lower vaccination rates could result in a nationwide outbreak.

Reading the US news media over the past two weeks, you might

conclude that that day has come. The current US measles outbreak, which began in December and was first reported in late January, has blown up into a national debate over the rights of parents to decide whether their children should be vaccinated. But by global standards, it is a tempest in a teapot: as of 6 February, measles had struck 121 people in 17 states and the District of Columbia.

Those numbers are unremarkable. Since October, a measles outbreak has affected more than 370 people in Germany; it saw almost 1,800 cases in 2013 and more than 1,600 in 2011. The Philippines had more than 50,000 cases in 2014. The United Kingdom had only 137 cases last year, but in both 2012 and 2013 it had close to 2,000 (see page 148).

In fact, even by US standards, the current outbreak is not unprecedented. Last year, a much larger outbreak was sparked by Amish missionaries returning from the Philippines to Ohio, where low vaccination rates among the community caused 383 cases.

Perhaps that incident stayed out of the national spotlight because it was an unusual set of circumstances that occurred in an isolated rural community. But the current outbreak centres on 'the happiest place on Earth' — Disneyland in southern California. At least 42 people seem to have been exposed to measles at the theme park, which receives an estimated 16 million visits a year.

Fortunately for the public's health, attention around the outbreak has come down in favour of vaccination and against the myths about its dangers. Public opinion has turned against parents and physicians who are suspicious of vaccines. Two potential Republican presidential candidates, Governor Chris Christie of New Jersey and Senator Rand Paul of Kentucky, at first declared that parents should have the right to decide whether their children are vaccinated, and then had to clarify their positions in the face of harsh criticism.

Whether or not the theme park's involvement in the episode contributed to the media coverage, Disneyland's cherished place in US culture makes it ideal for an infectious-disease outbreak. It is popular with international tourists eager for a quintessential American experience, who as

a group are less likely than US residents to be vaccinated. The park also hosts large numbers of infants less than one year old — younger than the age at which the first measles shot is generally given in the United States.

And Disneyland is at the epicentre of the US anti-vaccine movement. Although 94.7% of US children entering school at around age 5 are vaccinated against measles, in hundreds of California schools the percentage of vaccinated children falls well short of the 92%

"Disneyland is at the epicentre of the US anti-vaccine movement."

considered necessary to produce the 'herd immunity' that prevents transmission of the disease. The state's public-health department reports that 2.54% of children entered school in 2014 with an exemption from vaccination based on personal belief.

The federal government has little say in who gets a measles shot — those rules are written by individual states. Most, like California, allow parents to send their children to school unvaccinated by claiming a religious or philosophical objection to the practice. But two — Mississippi and West Virginia — allow only medical exceptions. And that, many observers have argued, is why Mississippi, one of the poorest states in the union, has the highest percentage of 5-year-old children who have received vaccination for measles, mumps and rubella.

Last month, the Mississippi state legislature was considering a bill to allow the same types of personal-belief exemption that most other states allow. But on 3 February, a committee in the state's House of Representatives killed the proposal. On 4 February, legislators in California said that they would introduce a bill to adopt the same strict rules as Mississippi. And several other states, including Maine, Minnesota and Oregon, are considering measures that would require parents to consult with a physician about vaccines before being granted an exemption.

That is a step in the right direction. Parents, of course, have the right to decide what is best for their children. But when it comes to vaccination, those decisions should be based on complete and accurate information about the risks and benefits. ■

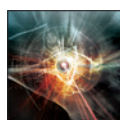
A single light

A year of illumination switches on with a Nature special issue.

Among the measures approved at the 68th session of the United Nations General Assembly in December 2013 were resolutions to develop "a world against violence and violent extremism" and "measures to eliminate international terrorism". Against such targets, the goal of UN resolution A/RES/68/221, passed in the same session, might seem unambitious: to recognize the importance of light in the lives of the citizens of the world.

Some 42 days into that effort — officially called the International Year of Light and Light-based Technologies 2015 — *Nature* is doing its bit. In this special issue, we offer a series of articles that explore how researchers are pushing the properties of light to new extremes, and the impact that these studies are already having and could have in future. The print-journal package begins on page 153 and there is more available online at nature.com/light2015.

Light and science have been entwined for more than a thousand years; light and life for much longer. This is reflected in the goals of the UN celebration, from discussions of solar energy and its crucial potential in tackling energy and climate problems to the societal impact of artificial light in our cities and homes, and how it guides development. According to



LIGHT
A *Nature* special issue
nature.com/light2015

the UN, "the 21st century will depend as much on photonics as the 20th century depended on electronics". If so, then more of the work that researchers are engaged in to understand and harness light — to make light work — will need to move out of the laboratory.

Light inspires, too. The organizers of the UN year of light are seeking people to follow in the (chunky) footsteps of writers such as Fyodor Dostoyevsky and Johann Wolfgang von Goethe, who wrote about optics. They invite those who feel that they have something to say about light and any phenomena or feeling connected to it to enter a literary competition. Poems, short stories, essays and plays are welcome, but must be submitted by the end of next month (see go.nature.com/5pqdnt for details). Winning entries will appear in a special anthology — published a thousand years after Ibn al-Haytham's classic treatise, *Book of Optics* (see page 164).

Light in 2015 may be all about applications and technology, but it retains a powerful theoretical pull on the scientific mind. Countless children in night-time gardens have been astonished and intrigued by the news that the light arriving from distant stars is a historical record — the stars themselves could be long gone even as the light carries their image on its journey. Generations of students have tried to decipher whether light is a wave or a particle, and in doing so have come to accept that scientific reality demands a greater tolerance of uncertainty than the textbooks suggest. Albert Einstein's general theory of relativity — the centenary of which is recognized as part of the UN's celebration — has come to represent an intuition warped just as much as the light in the gravitational field that it describes.

Light has outgrown its metaphorical role as an answer to questions; light itself remains a puzzle. To solve that puzzle is an ambition that deserves the recognition that the coming months will shine on it. As the biochemist and author Isaac Asimov put it: "There is a single light of science, and to brighten it anywhere is to brighten it everywhere." ■



And the winner is: not science

Portrayals of science in the cinema are growing in sophistication — but not exactly at the speed of light, says **Colin Macilwain**.

Some Hollywood stardust will sprinkle across the world of science this week, in the build-up to next weekend's Oscars ceremony in Los Angeles. Uniquely in the history of the silver screen, two of the leading contenders for Best Picture concern the lives of two great scientists, mathematician Alan Turing and physicist Stephen Hawking.

Both films have been widely praised. But unfortunately, in terms of shedding light on what made these scientists tick — or furthering the art of film-making — *The Imitation Game* and *The Theory of Everything* each leave a great deal to be desired.

The two films were produced in the United Kingdom, not in Hollywood. But they each feature a catalogue of clichés, of eccentric scientists and true love, well worthy of Hollywood in its gory heyday.

You may say that it is too much to ask — but I think scientists deserve to see major, fact-based feature films about science present their lives in ways that resonate, at least to some extent, with the world of science as it really is. Most of us can recognize the authentic when we see it; in the case of these two films, we don't.

It is ironic that although Hollywood has shown itself capable of producing, on occasion, complex, postmodern masterpieces such as 2004's *Crash*, film-makers here in the United Kingdom still churn out the sort of sentimental slop that British satirists used to make a living by sending up, a quarter of a century ago. (I refer younger and non-UK readers to the genius of the *Comic Strip* series.)

The Imitation Game, Morten Tyldum's portrayal of Alan Turing, is the greater disappointment of the two. Benedict Cumberbatch's performance as Turing has been widely — and justifiably — lauded. But the script, unfortunately, portrays Turing as a dysfunctional, almost autistic, individual and trots through clichés of how a 'genius' treats his peers with all the finesse of a children's fable.

All we learn about the project to bust the German Enigma cipher in the Second World War is that everyone was doing it all wrong until our erstwhile, eccentric hero turns up, argues with everyone in sight and relentlessly ploughs his own furrow, whatever that may be (we are never told). The film is significantly weaker for saying almost nothing about the nature of the problem, or about Turing's role — relative to others, inside and outside the project's base at Bletchley Park — in the conception and implementation of what we now call the computer.

It also groundlessly alleges that Turing's homosexuality made him turn a blind eye to a likely spy at Bletchley Park — a piece of worthless and defamatory melodrama that seems gratuitous, given the ample material provided by Turing's real life story.

Greater emotional nourishment, at least, is

forthcoming from *The Theory of Everything*, in which Eddie Redmayne skilfully carries the viewer into the world of Stephen Hawking, as his body is progressively ravaged by motor neuron disease.

Hawking is portrayed sympathetically but convincingly, and the film addresses the great issues of his life outside science — the impossible demands placed on his first wife, Jane, on whose memoir, *Travelling to Infinity: My Life with Stephen* (Alma, 2008), the film is largely based, and the lack of support offered to the couple from the outside world.

Some critics have said that the film ought to have been even harsher. The book on which it is based is a softer version of Jane Hawking's earlier memoir, *Music to Move the Stars* (Macmillan, 1999), now out of print. (Intriguingly, second-hand copies are trading on Amazon for several hundred pounds.)

I enjoyed and believed this film — but it makes only a cursory effort to describe or address Hawking's scientific trajectory. Given his status as perhaps the world's best-known living scientist, there is something unsettling about that.

Both films present a bombastic, simplistic and 'hero-takes-all' picture of science — a picture that is still promoted heartily through the Nobel prizes, and by much science writing.

I prefer the more jaundiced view taken by Paul King's family film *Paddington*, in which geographer Montgomery Clyde is expelled from his learned society for failing to kill and bring back bears that he has found in Peru.

As has been widely noted, both audiences and critical attention have been shifting from cinema to the smaller screen, as television writers adapt to a twenty-first century in which people are growing wise to the clichés foisted on them in the past.

A more-nuanced approach to storytelling has emerged in countless television series, from *Breaking Bad* to *House of Cards*. None of these, so far, is built around the world of science, but a similar intelligence shines through the world-beating science-based sitcom, *The Big Bang Theory*. Trite as some of its scripts may be, *Big Bang* has a stronger grasp than either of these movies of how science really works, bouncing along on a melee of inspiration, treachery, serendipity and teamwork.

Big Bang's barrage of cameos, from the likes of physicist Brian Greene and even Hawking himself, speaks to its credibility and fan base inside the scientific community. Its appeal carries an important message, too: scientists are not circus freaks; they are just people, whose work lets them express their inner nerd. It would be nice to see something about science on the big screen that carried half as much conviction. ■

Colin Macilwain writes about science policy from Edinburgh, UK.
e-mail: cfmworldview@googlemail.com

**MOST OF US CAN
RECOGNIZE THE
AUTHENTIC
WHEN WE SEE IT;
IN THE CASE OF THESE
TWO FILMS,
WE DON'T.**

➔ **NATURE.COM**
Discuss this article
online at:
go.nature.com/6akuhq

RESEARCH HIGHLIGHTS

Selections from the
scientific literature

CHEMISTRY

Nanoparticles stuck on tape

Researchers have found an easy way to deposit metal nanoparticles on a surface — using sticky tape.

Adding nanoparticles to a surface can give it properties such as electrical conductivity. Bartosz Grzybowski at Northwestern University in Evanston, Illinois, and his colleagues showed that when commercial Scotch tape is peeled away, bonds within the tape polymer break and radicals form on its surface. These then react with metal salts to produce metal nanoparticles on the tape.

When the team placed peeled tape into a solution of silver nitrate for several hours, the tape turned yellow-orange — indicating that silver nanoparticles had formed. The silver-coated tape showed antibacterial activity and remained sticky.

J. Am. Chem. Soc. <http://doi.org/zzn> (2015)



MATERIALS

Capsules collect carbon dioxide

Microcapsules containing a liquid carbonate solvent could capture carbon dioxide from power plants more efficiently than existing methods.

Currently, CO₂ is captured at power plants by passing the flue gas over a solution of liquid monoethanolamine. The liquid is corrosive, forms toxic by-products and must be heated to high temperatures to recover the CO₂ and regenerate the solvent. Jennifer Lewis and her colleagues at Harvard University in Cambridge,

Massachusetts, created microcapsules made of a highly porous silicone skin containing a carbonate solvent. These solvents absorb CO₂ slowly, but encapsulation of solvent boosts the absorption rate tenfold (compared to pools of liquid carbonate) by increasing the surface area.

The capsules (pictured) are chemically stable and environmentally benign, and CO₂ can be recovered by modest heating.

Nature Commun. **6**, 6124 (2015)

STEM CELLS

Injected cells fix brain injury

Cells derived from human stem cells repair brain damage in irradiated rats, suggesting a possible therapy for survivors of brain cancer.

Radiation treatment of brain cancer can impair memory, attention and learning. Viviane Tabar at the Memorial Sloan Kettering Cancer Center in New York and her colleagues used human embryonic stem cells to make progenitor cells that form oligodendrocytes, which insulate nerve fibres, boosting the speed of electrical impulses. The team injected these cells into the brains of rats that had been exposed

to radiation. The animals did better at learning and memory tasks than irradiated rats that had not received cells, and about as well as untreated rats.

Analysis of rat brain tissue revealed that the transplanted cells re-insulated nerves in many parts of the brain.

Cell Stem Cell **16**, 198–210 (2015)

ECOLOGY

Bee behaviour sees colonies collapse

Honeybee colonies could be collapsing because younger bees are flying out to forage, raising their risk of death.

Many bee colonies are failing, probably because of parasites, pathogens and

pesticides. Bees react to such stressors by foraging at a younger age, so to learn how this might cause rapid population declines, Andrew Barron at Macquarie University in Sydney and his colleagues radio-tagged bees in experimental colonies to monitor their flight behaviour. The insects that began foraging earlier in life completed fewer successful trips and had a lower survival rate than those that foraged at the normal age.

Mathematical models showed that the resulting decrease in food for the colony and the increased forager mortality over time led to rapid colony collapse. The authors suggest that supplemental feeding of colonies could help

to stave off bee declines.

Proc. Natl Acad. Sci. USA <http://dx.doi.org/10.1073/pnas.1422089112> (2015)

CLIMATE CHANGE

Aerosols reduce Arctic warming

Particles suspended in the atmosphere have decreased the amount of warming caused by greenhouse gases in the Arctic, but this could change as future air pollution is reduced.

Aerosols have a cooling effect by reflecting sunlight back into space. Mohammad Reza Najafi at the University of Victoria in Canada and his colleagues analysed nine climate models running from

JOHN VERICELLA, ROGER AINES (LLNL); JENNIFER LEWIS (HARVARD)

1913 to 2012, comparing simulations with and without greenhouse gases, aerosols and other climate drivers. Their results show that aerosols have offset 1.3–2.2°C of Arctic warming from greenhouse gases, limiting the observed warming to 1.2°C. With aerosol emissions projected to drop in the coming decades, the rate of the warming is likely to increase.

The team says that its results underscore the reliability of the climate models, which simulate 8.3°C of warming in the Arctic in a high-emissions scenario by the end of the century. *Nature Clim. Change* <http://dx.doi.org/10.1038/nclimate2524> (2015)

ANIMAL BEHAVIOUR

Chimps learn new calls for food

Captive chimpanzees learn new grunts from neighbours to refer to foods — the first evidence of such behaviour in non-humans, according to a study.

To see whether chimps (pictured) show flexibility in the calls they use to refer to everyday objects, Simon Townsend at the University of Zurich, Switzerland, and his team compared the grunts of seven chimps that were moved from a safari park in the Netherlands to join six chimps in a UK zoo. A year after the move, the Dutch chimps referred to apples with a high-pitched call, in contrast to the deep-timbred grunts of the UK chimps. But after three years, the Dutch chimps had adopted their neighbours' calls.

The findings suggest that social learning of referential words in humans could have

a longer evolutionary history than was thought. *Curr. Biol.* <http://doi.org/zzd> (2015)

NEUROSCIENCE

Deep-brain zap for addiction

An electric current sent deep into the brain, together with a therapeutic drug, can reverse the symptoms of cocaine addiction in mice.

Christian Lüscher at the University of Geneva in Switzerland and his colleagues implanted an electrode into the brains of cocaine-addicted mice. Stimulating the animals' neurons at a low frequency only temporarily relieved symptoms of addiction after the mice were injected with cocaine. But when the researchers also gave the animals a drug that blocks receptors for the neurotransmitter dopamine — involved in addiction and reward — the symptoms abated for longer. Neural connections that were overactive because of cocaine exposure also functioned normally again.

The researchers say that this approach could be a potential therapy for humans with addiction and other neural disorders.

Science 347, 659–664 (2015)

ANTHROPOLOGY

Soggy climates affect language

In warm, moist climates, human languages developed with more complex linguistic tones than did those in colder, drier regions.

It is thought that language is not influenced by ecological factors. However, Caleb Everett at the University of Miami in Coral Gables, Florida, and his colleagues concluded the opposite after looking at studies of vocal-cord biology and comparing the geographic origins of more than 3,700 languages with the humidity and annual average temperatures of those regions. The vocal-cord data showed

SOCIAL SELECTION

Popular articles on social media

Lab size matters for productivity

To publish the most papers, labs should ideally have 10 to 15 members, according to a much-discussed study in *PeerJ PrePrints*. Adding more graduate students and postdocs beyond that number does not guarantee a continued rise in high-impact papers, the study found, partly because the extra workers tend to be much less productive than the principal investigator (PI). Mark Pallen, who heads a microbiology lab at the University of Warwick, UK, tweeted “Nice that PIs matter!” But Jessica Chong, a geneticist and postdoc at the University of Washington in Seattle, called it an “odd analysis” on Twitter, adding, “we expect PIs to ‘produce’ more papers than any other lab member. They’re authors on all papers!” *PeerJ Prepr.* 3, e812v1 (2015)



Based on data from altmetric.com. Altmetric is supported by Macmillan Science and Education, which owns Nature Publishing Group.

➔ **NATURE.COM**
For more on popular papers:
go.nature.com/kbhjkg

that cords produce sounds of varying pitch more accurately in moist air than in dry air. Languages with complex tone, such as Mandarin Chinese, originated mainly in warm, moist climates, whereas languages such as English, which have little or no tone, came from arid or cold regions.

Proc. Natl Acad. Sci. USA 112, 1322–1327 (2015)

EVOLUTION

A hint at how hearing evolved

Early four-legged vertebrates may have been able to hear sounds on land, even though they lacked key ear structures.

Christian Christensen at Aarhus University in Denmark and his colleagues studied the hearing of the African lungfish (*Protopterus annectens*; pictured), the closest living relative of early tetrapods that began moving onto land around 350 million years ago.

The middle ear, which senses changes in air pressure caused by sound, is missing in lungfish. The researchers found that low-frequency sounds in air caused the lungfish's head to



vibrate. Its brain responded to these frequencies, suggesting that the animal detects airborne sounds using the vibrations.

Another study by some of the same authors looked at salamanders, which also lack middle ear structures and live in water and on land. The team showed that even juvenile salamanders, which are fully aquatic, can detect sound in air.

The findings suggest that early tetrapods were pre-equipped to hear sounds in air, which probably helped them to adapt to life on land and eventually led to more-modern middle ear structures.

J. Exp. Biol. 218, 381–387 (2015); *Proc R. Soc. B* 282, 20141943 (2015)

➔ **NATURE.COM**
For the latest research published by Nature visit:
www.nature.com/latestresearch



TOM MCHUGH/SPL

FLORIAN WOELLERS/RZSS

SEVEN DAYS

The news in brief

POLICY

US security

US President Barack Obama has incorporated the dangers posed by global warming and poverty into a new national-security strategy. Released on 6 February, the document says that both economic and environmental threats can hamper growth, foster extremism and ultimately lead to military interventions. The document also highlights the role of scientific and technological innovation in promoting cleaner domestic energy, stating that the United States “can and will lead the global economy while reducing our emissions”. The White House’s previous national-security strategy was outlined in 2010.

Three-person IVF

The UK House of Commons voted on 3 February to legalize a gene-therapy technique that could help women to avoid passing genetic defects on to their children through mutations in their mitochondria — the cell’s energy-producing structures. The technique, known as mitochondrial replacement, uses healthy donated eggs instead of the mother’s diseased eggs to create ‘three-person embryos’ for *in vitro* fertilization (IVF). The UK House of Lords must also approve the measure, which would authorize the country’s fertility regulator to allow mitochondrial replacement in the future. See page 145 for more.

Bee petition

A coalition of 11 US environmental groups has urged President Barack Obama to significantly toughen rules governing the use of neonicotinoid pesticides, which have been linked to

declines in bee numbers. In a letter to the White House, groups including the National Audubon Society, the Sierra Club and Friends of the Earth called for more research on the impact of these chemicals on pollinators. They also propose a ban on treating seeds with neonicotinoids because this affects the entire plant. A federal strategy for dealing with bee health is expected in the coming months.

RESEARCH

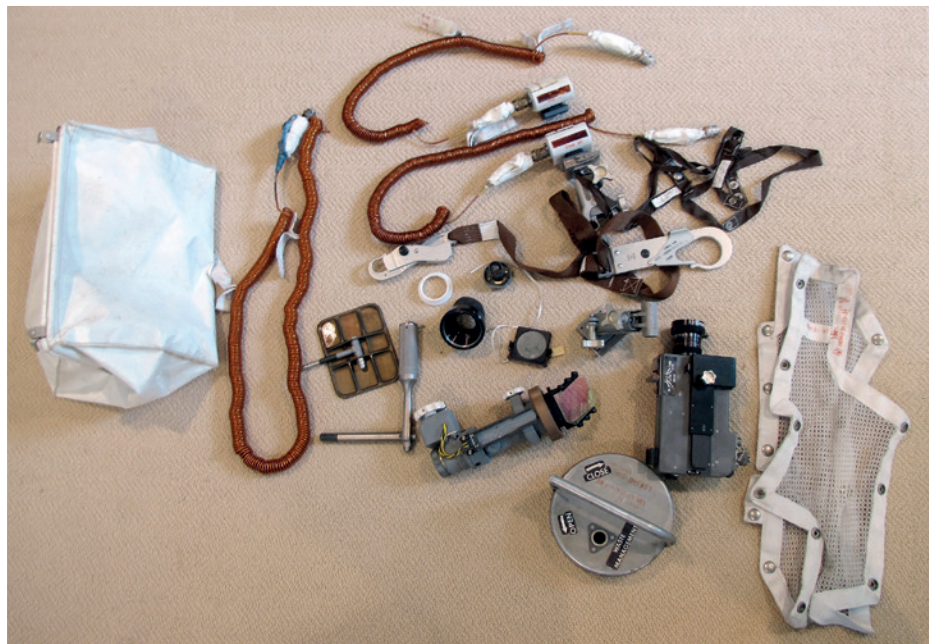
Clean-coal cut

The US Department of Energy has pulled out of its flagship clean-coal project, dubbed

FutureGen, more than a decade after it was proposed. The US\$1.7-billion demonstration project, modified in 2010, was intended to provide climate-friendly power by retrofitting a coal-fired power plant in Illinois to capture carbon dioxide and pipe it into a saline aquifer some 1,200 metres below ground. The department had committed \$1 billion to the project, but pulled out owing to ongoing questions about private investments. By law, the federal money had to be spent by September this year. See go.nature.com/2pymmo for more.

Japan Venus probe

Japan’s errant Akatsuki probe is to get another chance at studying the meteorology of Venus after a failed attempt in 2010. The Japan Aerospace Exploration Agency announced on 6 February that it will try to insert the ¥25-billion (US\$211-million) craft into Venus’s orbit this December. If this is successful, Akatsuki will use remote sensing to observe the planet’s clouds, atmosphere, lightning and surface conditions, allowing a comparison with similar mechanisms on Earth. The probe has been orbiting the Sun since December 2010, after a malfunctioning



NATIONAL AIR AND SPACE MUSEUM

Neil Armstrong’s Moon bag revealed

A white bag used by Neil Armstrong during his 1969 Moon landing was made public last week. It was found in a cupboard by Armstrong’s widow, Carol, after his death in 2012. The contents (pictured), including a camera and waist tether, were analysed by curators at the Smithsonian National Air and Space Museum

in Washington DC, who concluded that the items were from the *Eagle* lunar module (LM). Transcripts of Armstrong speaking to astronaut Michael Collins helped to confirm the bag’s authenticity: “That one’s just a bunch of trash that we want to take back — LM parts, odds and ends, and it won’t stay closed by itself.”

thruster stopped it decelerating enough to drop into the planet's orbit.

Aquarium census

The Shedd Aquarium in Chicago, Illinois, will take the first census of microbes that live in aquariums. As part of the Aquarium Microbiome Project, which launched on 3 February, researchers from Shedd and partner institutions will analyse how the microbes living in the aquarium's tanks and on resident sea animals differ from those in natural aquatic environments. The project will also assess the impact of pollutants on microbial ecosystems, and plans to release its first data later this year.

PEOPLE

FDA chief quits

The commissioner of the US Food and Drug Administration (FDA) announced her resignation on 5 February. Margaret Hamburg (pictured) has led the regulatory agency for nearly six years, during which time it established programmes to speed up drug approvals, laid the groundwork to regulate electronic cigarettes, and proposed guidelines for regulating medical-diagnostic tests. Hamburg will continue in the post until the end of March. The FDA's chief scientist,



Stephen Ostroff, will serve as acting commissioner, but the agency is yet to announce a permanent replacement.

Physicist dies

Val Fitch, the physicist who, with James Cronin, discovered a fundamental asymmetry between matter and antimatter, died on 5 February, aged 91. In 1964, while both were at Princeton University in New Jersey, Fitch and Cronin showed that particles of antimatter do not simply behave as mirror-symmetry counterparts of matter particles. This violation of a law known as charge-parity symmetry is believed to be the reason that the Big Bang did not produce a Universe that is equal parts matter and antimatter. The two physicists were awarded the 1980 Nobel Prize in Physics for their discovery.

Climate libel

A Canadian climate scientist has been awarded Can\$50,000 (US\$40,200) in a libel lawsuit

against the *National Post* newspaper. Andrew Weaver, now also a politician for his province, challenged four *Post* articles published in 2009–10 in the wake of the 'Climategate' scandal, when hacked e-mails from UK climate scientists were made public. The articles called Weaver a "climate alarmist", and he said that they implied he was untrustworthy and unscientific. The 5 February ruling by the British Columbia Supreme Court says the *Post* should take down the articles and that they "adversely impact on Dr. Weaver's reputation and integrity as a scientist".

FUNDING

Disease initiative

The Global Health Innovative Technology Fund in Tokyo launched a programme on 5 February to spur the development of drugs, vaccines and diagnostic tools for infectious diseases prevalent in developing countries. The Grand Challenges initiative will invest ¥234 million (US\$2 million) per year in research into diseases such as malaria, tuberculosis and Chagas disease. The first grant-winners are expected to be announced in August. The fund was founded in April 2013 with money from the Bill & Melinda Gates Foundation

COMING UP

12–16 FEBRUARY

The American Association for the Advancement of Science's annual meeting takes place in San Jose, California. It will focus on how information and imaging technologies are transforming science. go.nature.com/kprsx8

14 FEBRUARY

The Rosetta craft is due for its closest encounter yet with comet 67P/Churyumov–Gerasimenko. The probe will swoop just 6 kilometres from the surface, and with the Sun at its back should get the first shadow-free images of the comet. But it will not be specifically searching for its lost partner — the lander Philae.

in Seattle, Washington, the Japanese government and six Japanese pharmaceutical companies.

UK science funding

The United Kingdom's science academies are calling for a huge hike in research spending from whichever political party wins the country's election in May. Representatives of the British Academy, the Royal Academy of Engineering, the Royal Society and the Academy of Medical Sciences said on 10 February that the next government should aim to spend 3% of gross domestic product (GDP) on research and development by 2020, with 1% of this coming from the public purse. The current UK research spend is 1.73% of GDP, of which 0.5% is from the public sector. Government science spending has declined in real terms since 2010.

➔ NATURE.COM

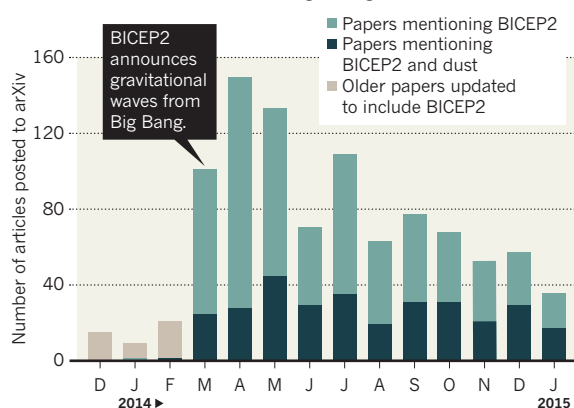
For daily news updates see: www.nature.com/news

TREND WATCH

Manuscripts posted to the preprint server arXiv show how cosmologists rapidly embraced, then began to doubt and lose interest in, one of 2014's most sensational announcements: the discovery of gravitational waves from the birth of the Universe. Soon after the March announcement, papers questioning the result started to emerge. The final nail in the coffin came last month, when researchers conceded that dust in the Milky Way accounted for the signal seen by the telescope BICEP2.

THE TRAIL OF DUST

Interest waned in the BICEP2 experiment as it became clearer that dust had been mistaken for a signal of gravitational waves.

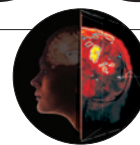


NEWS IN FOCUS

EMERITUS GRANTS 'Cash for retirement' idea draws ire **p.146**

MEASLES Inside the race to eradicate a killer disease **p.148**

TEXT-MINING Words that triumphed in UK research audit **p.150**



OPTICS Physicists find ways to see through walls **p.158**

TED HOROWITZ/COREIS



Three-person *in vitro* fertilization prevents women from passing on potentially harmful mutations in mitochondrial DNA.

REPRODUCTIVE BIOLOGY

World hails embryo vote

UK move to allow pioneering fertility technique could spur other countries to relax rules too.

BY EWEN CALLAWAY

Following a 3 February vote in the UK House of Commons, the world may once again look to Britain to lead in fertility treatments, 37 years after *in vitro* fertilization (IVF) was pioneered in the country.

The vote lifts a ban on gene-altering fertilization techniques known as mitochondrial replacement, or three-person IVF, in which mitochondria — the cell's energy-processing structures — from a donor's egg cell contribute to a couple's embryo. The procedures are intended to prevent the transmission of diseases caused by mutations in mitochondrial

DNA. The vote, won by 382 in favour versus 128 against, will still need to be confirmed by the House of Lords, which is widely expected to pass the law. Once approved, the Human Fertilisation and Embryology Authority (HFEA), Britain's fertility regulator, will be allowed to license clinics to carry out the procedures from October, although it could be some time before the first human trials begin.

Many reproductive biologists see this as a step that will affect the field on a global scale. "We've been hoping that the UK will take the lead," says Shoukhrat Mitalipov, a stem-cell scientist at Oregon Health & Science University in Portland. His team hopes to apply to

the US Food and Drug Administration (FDA) for permission to conduct clinical trials of mitochondrial replacement. Although its regulatory debate is a bit behind the United Kingdom's, "the US is going down the same path", Mitalipov says.

An estimated 1 in 5,000 children are born with diseases caused by mitochondrial mutations, which typically affect energy-hungry tissues such as the brain, heart and muscles. All mitochondrial DNA is inherited from the mother, and some women carry harmful mitochondrial mutations without having symptoms themselves. Their children can experience debilitating and sometimes ▶

► fatal conditions such as muscular dystrophy or heart disorders.

Before giving a green light to clinics wishing to offer the treatment, the HFEA will probably want further evidence that the procedure is safe, and will vet applications on a case-by-case basis.

The United Kingdom was one of the few countries that explicitly banned mitochondrial replacement by law. In many countries, including China and Japan, the techniques are prevented by regulations that should be simpler to overturn, without legislative intervention, says Tetsuya Ishii, a bioethicist at Hokkaido University in Japan.

The same applies to the United States.

Since 2001, the FDA has enforced a moratorium on mitochondrial replacement, after a New Jersey fertility clinic conducted a related procedure to improve conception rates. Mitalipov's push to launch a clinical trial of mitochondrial replacement set off a series of scientific, ethical and policy reviews that are still under way.

In February 2014, an FDA advisory panel held a two-day meeting to consider the science of mitochondrial replacement. The panel identified areas in which it wanted to see more data, such as the long-term health of monkeys conceived through the procedures, before it could move to allow mitochondrial replacement.

It will probably take two to five years to fill in these gaps, says Evan Snyder, a stem-cell biologist at Sanford-Burnham Medical Research Institute in La Jolla, California, who is chair of the FDA panel.

Australia is also pondering three-person IVF. Although the country's lawmakers opted against relaxing the rules after a 2011 review of human-cloning legislation, the UK vote will "provide enormous ammunition" for those seeking changes, says David Thorburn, a geneticist at the University of Melbourne, Australia. Still, he adds, "my gut feeling is that it's unlikely to succeed until this has been done in practice in the UK". ■

POLICY

NIH ponders 'emeritus grants'

A proposal to pay senior biomedical researchers to wind down their labs draws scepticism.

BY BOER DENG

For years, biomedical researchers in the United States have warned of a worrisome trend: as competition for grants increases, younger scientists are finding it harder to keep up. In 1980, the average age at which researchers received their first major award from the US National Institutes of Health (NIH) was 38; by 2013, this had risen to more than 45 (ref. 1). And the overall share of grant funding won by scientists younger than 36 withered from 5.6% in 1980 to just 1.2% in 2012.

Like ageing Crown princes, junior biomedical researchers in the United States face long years as leaders-in-waiting. Now, in a 3 February posting, the NIH has asked researchers whether the agency would be wise to give 'emeritus grants' to senior scientists to induce them to wrap up their research. The funding would "help to ensure the orderly transition of an experienced researcher's work when they wish to go on to something else, and also to recognize their legacy", says Sally Rockey, the NIH's deputy director for extramural research. If entrenched grant recipients leave the lab, the NIH hopes, more money will be available for early-career scientists.

Those who support the idea say that it could ease the pressure on senior researchers to continue working in order to bolster their retirement accounts, which in the United States largely depend on employee contributions. The evidence for this is anecdotal, however, and proponents of emeritus grants admit that few senior researchers complain that they lack money to close their labs.

But judging from more than 100 comments left on Rockey's widely read blog, many

researchers are highly sceptical of the plan, and are incensed by what they perceive as a retirement bonus for the already better-resourced. "The idea of allocating precious limited federal research dollars to a special 'emeritus' award appears, at best, tone deaf, and at worst, suggests underlying biases within the NIH that favour established researchers," says neuroscientist Benjamin Saunders, a postdoctoral researcher at Johns Hopkins University in Baltimore, Maryland.

Economic research suggests that paying older scientists to abandon their labs is unlikely to be the most effective way for the NIH to achieve its ultimate goal. Policies for adjusting markets work better when they are direct, says labour economist Richard Freeman of the National Bureau of Economic Research in Cambridge, Massachusetts. "If your goal is to have more young researchers have independent awards and positions, it would be more

efficient just to give them that," he says. "Any time you try indirect methods, there is much more uncertainty as to what will happen."

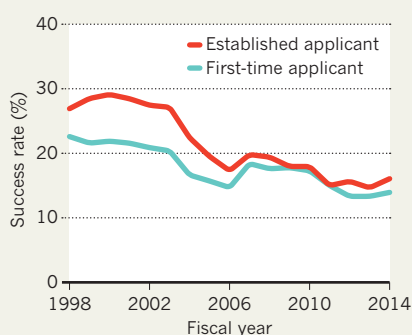
But the NIH has had mixed success with policies designed to give more money to new investigators. Since 2007, the percentage of grants won by new applicants has approached the share reaped by experienced scientists (see 'Age gap'), but critics say that funded proposals from younger researchers are of lower quality than those from older scientists. And despite the NIH's efforts, the average age at which a researcher wins his or her first award has not declined.

This may be partly because of broader demographic changes in the biomedical workforce. About 1 in 3 working scientists was over 50 in 2010, compared with 1 in 5 in 1993. This helps to explain why the average age of NIH principal investigators has risen. The age of first innovation itself might be increasing, too, according to analyses of patent filings and the age at which Nobel laureates win their prizes. Benjamin Jones, an economist at Northwestern University in Evanston, Illinois, has found that over the past century there has been a shift towards productive science at older ages, perhaps because innovation now requires more knowledge².

An even bigger challenge is an imbalance between the healthy supply of young scientists and the number of senior-level jobs, says Michael Teitelbaum, a demographer at Harvard Law School in Cambridge, Massachusetts. The problem has been exacerbated by erratic NIH funding. From 1998 to 2003, the agency's budget doubled, to US\$27.2 billion. Flush with grant money, academic research centres expanded, making jobs for biomedical-

AGE GAP

The US National Institutes of Health has sought to increase funding for new investigators, with mixed results.



SOURCE: NIH

science graduates plentiful and attracting more students to the field.

Fortunes subsequently reversed. Since 2003, the NIH's budget has contracted by around 25% in real terms, increasing competition for dwindling grant money among the surplus of early-career scientists created during the boom.

Without steady growth in the NIH budget, some have suggested that the solution is to train fewer graduates for careers in biomedical research. But the pipeline of new investigators shows no signs of drying up. In 2013, US universities conferred 8,471 biomedical PhDs. These joined thousands of other researchers eligible

that year for the NIH's Early Stage Investigator awards — 785 grants aimed at researchers who had graduated in the past decade. Too many heirs are awaiting too few crowns. ■

1. Daniels, R. J. *Proc. Natl Acad. Sci. USA* **112**, 313–318 (2015).
2. Jones, B. F. *Rev. Econ. Stat.* **92**, 1–14 (2010).

MARK MOFFETT/MINDEN/GETTY; RALPH LEE HOPKINS/GETTY



Ground finches (left) tend to have large beaks for cracking seeds, whereas warbler finches spear insects.

EVOLUTIONARY BIOLOGY

Darwin's finches join genome club

Scientists pinpoint genes behind famed beak variations.

BY GEOFF MARSH

Researchers have sequenced the genomes of all 15 species of Darwin's finches, revealing a key gene responsible for the diversity in the birds' beaks. The study, published online in *Nature* this week¹, also redraws the family tree of these iconic birds, whose facial variations helped Charles Darwin to formulate his theory of natural selection.

The finches are endemic to Ecuador's Galapagos archipelago and Costa Rica's Cocos Island. Their beaks are adapted to their preferred food: warbler finches, for example, spear insects with thin, sharp beaks, whereas ground finches crack open seeds with strong, blunter beaks. The birds are a textbook example of adaptive radiation, in which a single ancestor responds to a selective pressure — in this case, food availability — by diversifying into several species.

Darwin was the first to note this, during his

groundbreaking 1831–36 voyage aboard the HMS *Beagle*. “One might really fancy,” he wrote in his diary, “that from an original paucity of birds in this archipelago, one species had been taken and modified for different ends.” Almost two centuries later, his early suspicions have been widely confirmed.

Initially, the finches were classified on the basis of their physical characteristics. More recently, it has incorporated variations in key DNA sequences. But nobody had compared whole-genome data from all 15 species until a team led by Leif Andersson, a geneticist at Uppsala University in Sweden, analysed samples from 120 individual birds. “When we did the whole DNA sequence of all the species, we could redraw that tree,” he says.

Overall, the researchers found good agreement with current taxonomy, but also some interesting deviations. For example, they conclude that the ground finch *Geospiza difficilis*,

which is spread across six islands, actually comprises three species.

Andersson's team also discovered extensive mixing of genes between species. This is in line with field observations of hybrid birds made by study co-authors Peter and Rosemary Grant, evolutionary biologists at Princeton University in New Jersey who have worked in the Galapagos for decades. The genomic data reveal that the birds have been crossbreeding throughout their evolutionary history.

Darwin famously sketched his initial idea of phylogeny as a branching tree, above which he wrote “I think”. Now, says Peter Grant, “he might wish to redraw that tree by making connections between some of the branches, representing the hybridization and gene exchange”.

By looking at closely related finches that have different beak shapes, the researchers were able to pinpoint the genes responsible for beak morphology. One of those genes, *ALX1*, is involved in the facial development of vertebrates, including fish and mammals. In humans, for example, loss of *ALX1* leads to severe facial deformities².

In the finches, the gene displayed two distinct variants that matched up neatly with beak shape. Individuals from a species with a highly variable beak shape — the medium ground finch (*Geospiza fortis*) — had a mixture of the blunt and pointed gene variants. The finding dovetails nicely with work by the Grants that documents the species' rapid evolution as recently as the 1980s, when a drought affected the bird's food supply and its beak started to become more pointed to accommodate a new diet³.

Andersson suspects that *ALX1* drove that adaptation, but others say the picture is more complicated. Beaks “differ in many parameters, not just being blunt or pointed”, says Ricardo Mallarino, an evolutionary biologist at Harvard University in Cambridge, Massachusetts. Functional studies of *ALX1* should help to reveal exactly what the gene controls, he says. His colleague, evolutionary biologist Arkhat Abzhanov, says that *ALX1* may be especially important for finches with very specialized beaks.

What would Darwin make of the findings? “We would have to give him a crash course in genetics,” Grant says. “But then he would be delighted. The results are entirely consistent with his ideas.” ■

1. Lamichhaney, S. *et al. Nature* <http://dx.doi.org/10.1038/nature14181> (2015).
2. Uz, E. *et al. Am. J. Hum. Genet.* **86**, 789–796 (2010).
3. Grant, P. R. & Grant, B. R. *40 Years of Evolution: Darwin's Finches on Daphne Major Island* (Princeton Univ. Press, 2014).

Measles

A race to eradication

The US media is abuzz after an outbreak of measles in Disneyland, but the disease will keep popping up until it is wiped out worldwide.

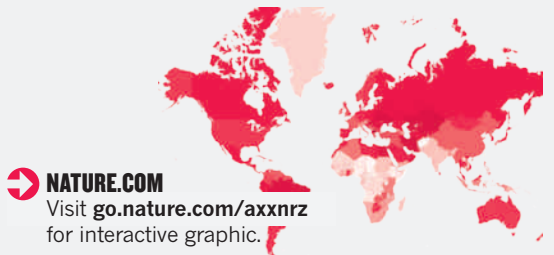
BY DECLAN BUTLER

Measles debate has reached fever pitch in the United States after an outbreak that began in December at Disneyland in southern California. Many media outlets and politicians have focused on the country's growing anti-vaccination movement. However, the bigger problem lies elsewhere. The United States was declared free of measles in 2000, and all outbreaks since then have been sparked by imported cases, which will continue to occur until measles is eradicated worldwide.

The World Health Organization (WHO) has set targets for 2015, but progress towards them has been slow (see 'Targets in trouble'). After the Measles & Rubella Initiative was founded in 2001, the number of cases and deaths fell. But progress against the disease started to stall in 2007 (see 'A fall ... then a stall') and vaccination coverage plateaued in 2010, when funding plummeted during the global economic slowdown (see 'Vaccination rising'). The WHO now concedes that few countries will attain anywhere near the targets.

The United States has some grounds for concern. Last year saw 644 cases in 27 states, a record high since 2000. And by 2013, the proportion of eligible children who had been vaccinated had dropped by 2% since 2004, to 91%. But the nation's vaccination coverage remains high compared with other countries (see 'Vaccination coverage worldwide'), and the number of cases is also small. The recent US outbreak infected 121 people. But China saw 107,000 people infected last year, and the Democratic Republic of the Congo (DRC) had 89,000 cases in 2013 (see 'Largest outbreaks'). The proportion of people who die varies depending on where you are (see 'Different conditions, different disease').

The last pockets of a disease are always the hardest to eliminate, as shown by polio eradication, which has been 'just around the corner' for years. But if the Measles & Rubella Initiative can get vaccinations back on track worldwide, measles may yet follow smallpox, the only killer human disease yet to be wiped out in the wild. ■

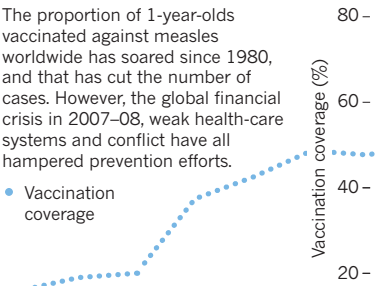


NATURE.COM
Visit go.nature.com/axnrz
for interactive graphic.

Vaccination rising

The proportion of 1-year-olds vaccinated against measles worldwide has soared since 1980, and that has cut the number of cases. However, the global financial crisis in 2007–08, weak health-care systems and conflict have all hampered prevention efforts.

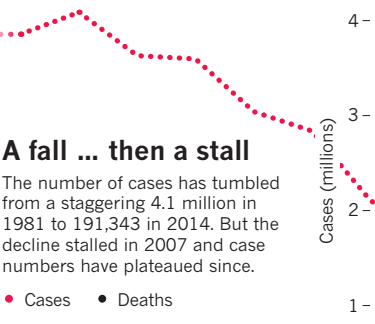
• Vaccination coverage



A fall ... then a stall

The number of cases has tumbled from a staggering 4.1 million in 1981 to 191,343 in 2014. But the decline stalled in 2007 and case numbers have plateaued since.

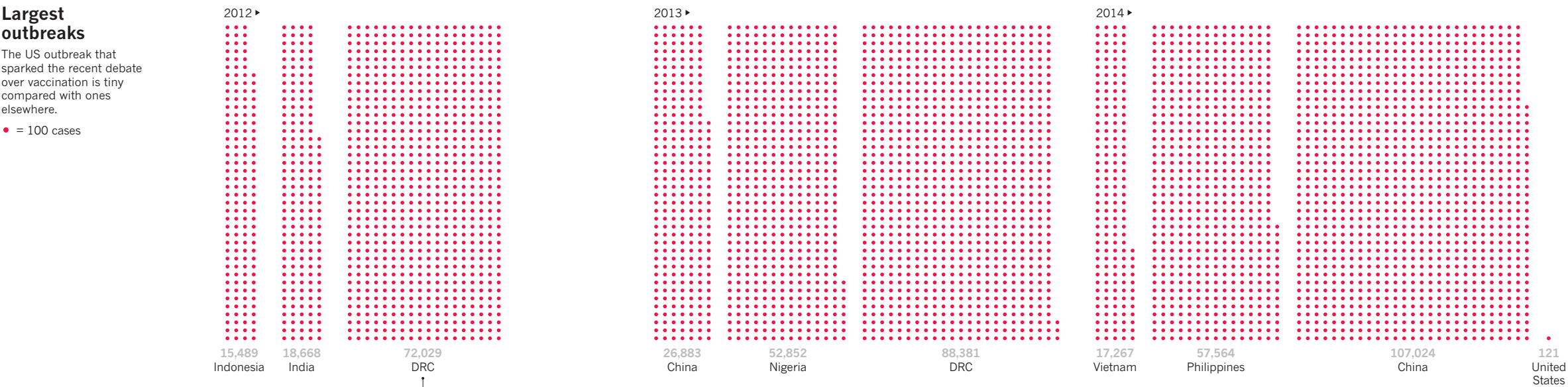
• Cases • Deaths



Largest outbreaks

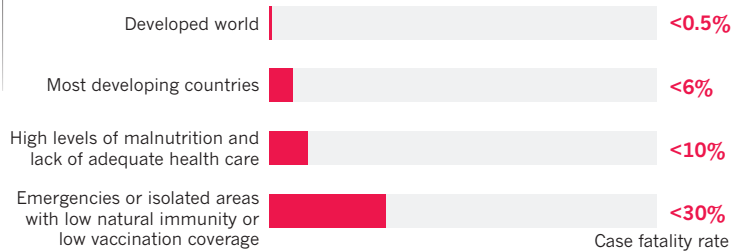
The US outbreak that sparked the recent debate over vaccination is tiny compared with ones elsewhere.

• = 100 cases

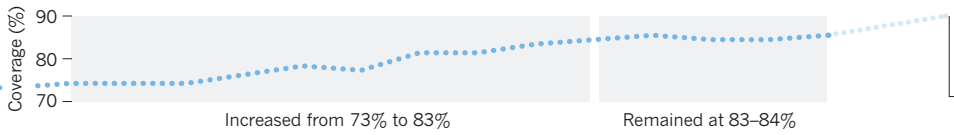


Different conditions, different disease

The proportion of infected people who die (case fatality rate) varies depending on many factors, including the quality of health care, nutrition and natural immunity.

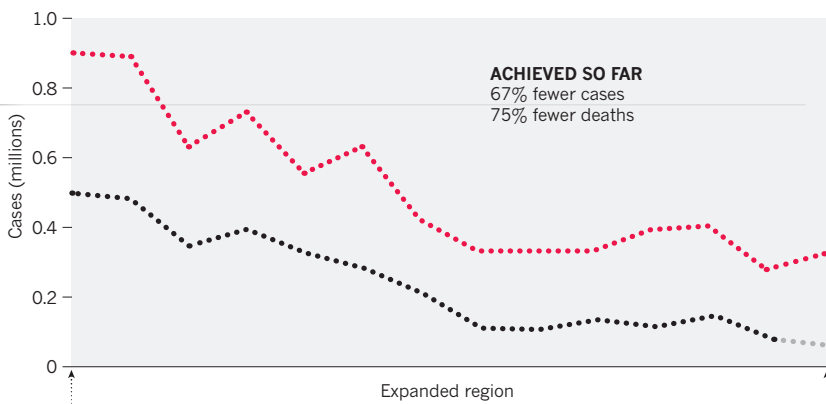


SOURCE: WHO. DESIGN: JASIEK KRZYSZTOFIAK/NATURE



Targets in trouble

The World Health Organization has set targets for 2015 that cut the number of cases and deaths and improve vaccination coverage from 2000 levels. But it looks unlikely that any of these will be met.



2015 VACCINATION TARGET
90% of children aged 1 year vaccinated.

2015 INCIDENCE TARGET
Fewer than 36,500 cases worldwide.

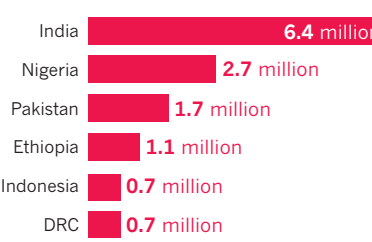
2015 MORTALITY TARGET
Fewer than 26,750 deaths.

Deadly data

The number of deaths worldwide has fallen and stalled in step with the number of cases. And at 145,700, the number of deaths in 2013 was slightly higher than that in 2007. The number of people measles kills depends on many factors (see 'Different conditions, different disease').

Vaccination lows

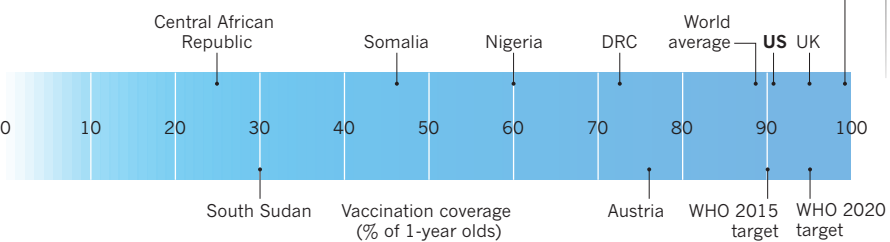
More than three-fifths of the estimated 21.5 million children who were not vaccinated against measles at 9 months of age in 2013 came from six countries.



Vaccination coverage worldwide

With 91% of 1-year-olds vaccinated, the United States comes in ahead of the WHO's 2015 target, but behind the 80 countries that have already attained the WHO's 2020 target of 95%.

36 countries, including Tanzania, Morocco, Greece, Cuba and South Korea, have attained 99% coverage.



Seven thousand stories capture impact of science

Language analysis reflects how case studies succeeded in a unique UK research assessment.

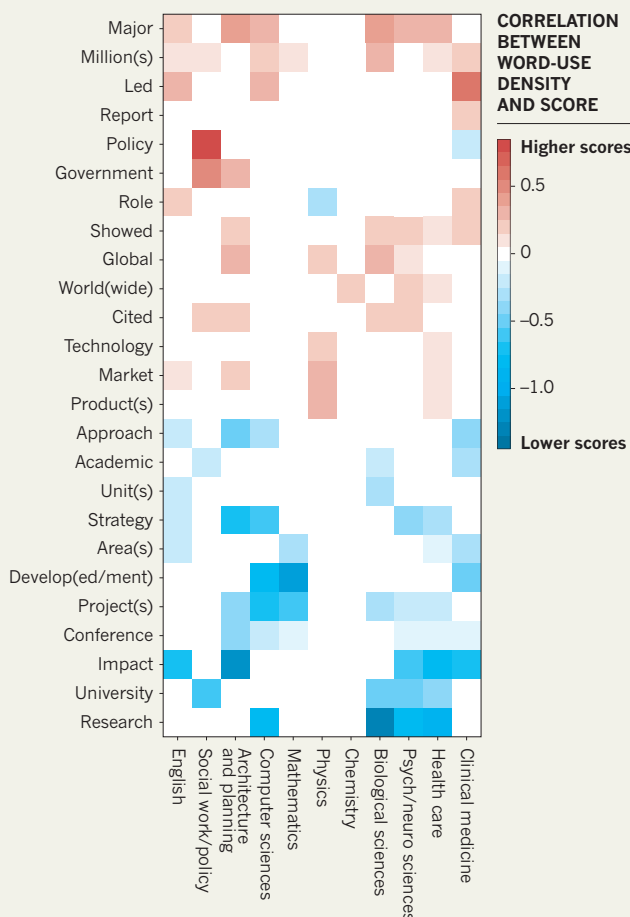
BY RICHARD VAN NOORDEN

Science benefits society in myriad ways — but how to identify and encourage work with high impact is an obsession of funding agencies the world over. Last month, the United Kingdom brought new data to bear on the problem: almost 7,000 case studies chronicling the economic, cultural and social benefits of the nation's scholarship, which were solicited as part of a unique assessment exercise. As policy-makers pore over the documents, *Nature* has commissioned its own analysis, revealing how researchers described the worth of their work to their paymasters, and hinting at buzzwords, including 'million' and 'market', that garnered high marks.

Many funding bodies ask academics to plan for the broader impacts of their work when they apply for grants. But the United Kingdom wanted to reward impact that had already been achieved, says Steven Hill, head of research policy at the Higher Education Funding Council for England (HEFCE). The country already has an audit culture: it grades the quality of university research every few years, and hands out £2 billion (US\$3 billion) annually on the basis of that assessment. For the 2014 audit, known as the Research Excellence Framework, or REF, HEFCE tweaked the rules. It added a requirement

POWER WORDS

In the UK REF assessment, impact case studies denser in terms such as 'major' and 'million' tended to be scored more highly, a text-mining analysis suggests.



REF = Research Excellence Framework. A value of +0.5 means that case studies mentioning a word at double the average density were correlated with an assessment score 0.5 higher than average. (In the assessment, scores ran from 1 to 4.) The words displayed are all high-frequency terms that showed statistically significant correlations in multiple disciplines. Eleven of 36 disciplines studied are shown. Case studies were analysed in clusters, because individual scores for each study were not released. For a larger image and methodology, see go.nature.com/jvcvcd.

that universities send in case studies detailing their work's wider impact during 2008–13, and announced that 20% of an institution's final grade would be based on these contributions (see *Nature* <http://doi.org/zx8>; 2014).

Meeting that challenge was a massive effort that sometimes involved hiring specialist writers and consultants. University College London alone wrote 300 case studies that took around 15 person-years of work, and hired four full-time staff members to help, says David Price, the university's vice-provost for research.

The results have impressed. "Every government wants to know the societal impact of its research," says Diana Hicks, who studies science and technology policy at the Georgia Institute of Technology in Atlanta. "The difficulty is how to do that broadly when you only have isolated case studies. Britain has cracked that problem and produced a wonderful data source."

The case-study narratives demonstrate "extraordinary breadth and depth," says Jonathan Grant, a public-policy researcher at King's College London. They range from chemists who used nanoparticles to prevent bacteria damaging the wood of a sunken sixteenth-century warship to economists who tested the effects of cash transfers to poor households in Mexico and Colombia.

SOURCE: PAUL GINSBURG/HEFCE

MORE ONLINE

EXPLAINER



Why the US nixed a carbon-capture project
go.nature.com/gqvgyx

MORE NEWS

- How light makes moths vulnerable to bats go.nature.com/wi2xza
- Tapeworms battle it out for dominance in host's guts go.nature.com/uxokp9
- Ice ages produce ocean-floor hills go.nature.com/olulms

NATURE PODCAST



Light-speed trading; sequencing Darwin's finches; and ancient Arabic optics. nature.com/nature/podcast

To draw further insights, *Nature* asked Paul Ginsparg, a physicist at Cornell University in Ithaca, New York, who has experience in text-mining, to run a statistical analysis of the language used in the case studies.

A straightforward word count revealed, unsurprisingly, that the terms ‘research’, and ‘impact’ were the most common, with 200,000 and 135,000 appearances respectively, after words such as ‘the’ or ‘and’ are removed. ‘Development’, ‘policy’ and ‘health’ also topped the lists. Notably, the documents name-check more than 190 countries, suggesting that the research has huge geographical reach.

Ginsparg also looked for statistically significant correlations between the use of certain words and the scores awarded. He found that across the disciplines, texts dense in words such as ‘million’, ‘market’, ‘government’, ‘major’ and ‘global’ tended to be given high scores by the judges, who were told to mark on the basis of ‘significance’ and ‘reach’ — whereas over-use of terms such as ‘conference’, ‘university’, ‘academic’ and ‘project’ correlated with lower grades (see ‘Power words’).

Although the correlations do not indicate causation, they might hint at judges’ preference for narratives of economic impact in particular, speculates Gemma Derrick, a researcher at Brunel University London who is examining how the studies were collected and assessed.

“I was sceptical about the ‘impact’ process, but now I think it’s a good thing,” says Price, who says it has revealed persuasive stories that the university can present to funders, industry partners, governments and alumni.

Some UK academics question whether the impact component to the research assessment will make a significant difference to how regional funders distribute their cash — and if not, whether it was worth adding. The formula that will link performance on the assessment

“Every government wants to know the societal impact of its research.”

to funding allocation will not be released until March, but it is already clear that universities that have traditionally excelled in the audit of academic output — Oxford, Cambridge and Imperial College London — also score highly on impact.

Internationally, some researchers criticize the idea of identifying research impact using case studies, rather than by tracking more quantifiable economic measures. “I am baffled why a scientific community would go through such a burdensome and artisanal system,” says Julia Lane, an economist at the American Institutes for Research in Washington DC and former director of a US government programme called STAR METRICS, which monitors the

economic benefits of money spent on research, including job creation, patents and spin-out companies. On 27 January, a network of European researchers — mainly economists — met in Brussels for the first formal meeting of an effort to trace how science funding in Europe leads to wealth and employment across society. The effort is strongly influenced by STAR METRICS.

Whether anyone will repeat the United Kingdom’s impact assessment remains an open question. “We know lots of other countries are interested in learning from our experience,” says Hill. Across the world, most countries that have introduced nationwide assessments of research quality, such as Australia and Italy, do not measure impact. Yet governments in both Sweden and the Czech Republic are currently considering an exercise similar to the REF.

Back in the United Kingdom, researchers are already preparing for the next performance audit, in 2020, with mixed feelings. “We will all be encouraged now to do more research that could form a case study — whether you think this is a good thing or not depends on your subject area,” says Dorothy Bishop, a neuropsychologist at the University of Oxford. “I have a concern that I may be stuck spending more time evaluating the impact of what I do and this will take me away from actually doing it.” ■ **SEE EDITORIAL P.137**

LIGHT FANTASTIC

Scientists are pushing the properties of light to new extremes. A special issue explores these frontiers.

From glorious rainbows to the intricate mechanics of the human eye, light lies at the heart of phenomena that have fascinated scientists for millennia. Today, the latest optical technologies — from lasers to solar cells — harness light to advance physics and to serve society's needs.

To put light itself in the spotlight, the United Nations designated 2015 the International Year of Light and Light-based Technologies. The celebration is also pegged to a string of anniversaries: Augustin-Jean Fresnel's proposal in 1815 that light is a wave; James Clerk Maxwell's 1865 electromagnetic theory; Albert Einstein's 1915 general theory of relativity; and in 1965, discovery of the cosmic microwave background (CMB) radiation and the development of optical fibres for communication.

Nature is paying its own tribute to light in this special issue. Contorting light is the goal of three physicists profiled in a News Feature on page 154: Miles Padgett twists laser beams to encode binary information; Pierre Berini reshapes light waves to speed up digital communications; and Margaret Murnane dissects X-rays into ultrafast attosecond pulses, one billionth of a billionth of a second long, to probe materials in exquisite details.

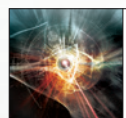
Some advances in the physics of light are of great benefit to biology and medicine. Borrowing from astronomers, biophysicists are developing techniques for seeing through opaque layers, by

detecting the minute glow of visible light scattered through body tissues. Such methods are likely to lead to more-powerful medical imaging, as explained on page 158.

In another sphere entirely, near-speed-of-light communications are set to transform financial trading as laser links between banking centres come online. But there are major risks, Mark Buchanan explains on page 161. Trading stocks in milliseconds pushes algorithms to their limits, exposing flaws that can escalate in seconds to cause hundred-million-dollar losses.

In a News & Views Forum on page 170, two cosmologists reflect on the clues to the origin of the Universe hidden in its oldest light, the CMB. And on page 164, physicist Jim Al-Khalili is dazzled by the afterglow of a 1,000-year-old treatise on the nature of light: Ibn al-Haytham's *Book of Optics*. An online collection will highlight key papers on light from journals across Nature Publishing Group throughout the year (see nature.com/yearoflight).

With so many facets, scientists' fascination with light looks unlikely to fade. ■

**LIGHT**

A *Nature* special issue
nature.com/light2015

LEADING LIGHTS

Shape it, squeeze it, energize it or tie it into knots. Scientists are taking light to new extremes.

| BY ELIZABETH GIBNEY |



| SHAPING LIGHT |
Miles Padgett

Physicist Miles Padgett starts to describe the concept of twisted light by taking down a rainbow-coloured spiral that hangs from the ceiling of his office at the University of Glasgow, UK. Then he stops and scours the room for more props: dinner plates, paper, pencils and even leftover Christmas chocolates.

Light is made of oscillating electric and magnetic fields, he explains. In a conventional laser beam, the oscillations are always in step, with the peaks and troughs lined up from one side of the beam to the other. (Padgett illustrates the flat, planar waves with a stack of dinner plates that he moves face-forward.)

But things get more interesting when parts of the beam fall out of step. This is where Padgett points to the spiral: the peaks of the wavefront can be manipulated to the point at which they curl around the beam's direction of motion in a corkscrew. This is twisted light, says Padgett, who has spent two decades learning to exploit its unique properties.

He has pioneered applications that range from moving cells without physically touching them to packing lots of information into an optical signal — and even tying light in knots. In the process, he has developed a rare instinct for the subject, say collaborators and colleagues. “Many other scientists might need to do a calculation, run a model or do an experiment before they can get an idea about how light should behave,” says Mark Dennis, a theoretical physicist at the University of Bristol, UK. “One of Miles’s great talents is having this knack at being able to anticipate what the results should be.”

Props are not the only thing in Padgett’s office. It houses the lab’s coffee machine, and doubles as its kitchen and common room — complete with sink. Padgett is a fan of productive chance encounters, and likes to keep the place buzzing with people picking each other’s brains.

It was a chance encounter that led him to twisted light in the first place. In 1994, as a research fellow at the University of St Andrews, UK, he had dinner with physicist Les Allen intending to discuss laser technology. But the conversation turned to Allen’s experiments with twisted light¹. Allen, then at the University of Essex in Colchester, UK, baited Padgett by saying that he knew how to give the light its twist using the stem of his wine glass as a lens. This strange idea had Padgett hooked. By 1997,

THE PHOTOGRAPHIC UNIT/UNIV. GLASGOW

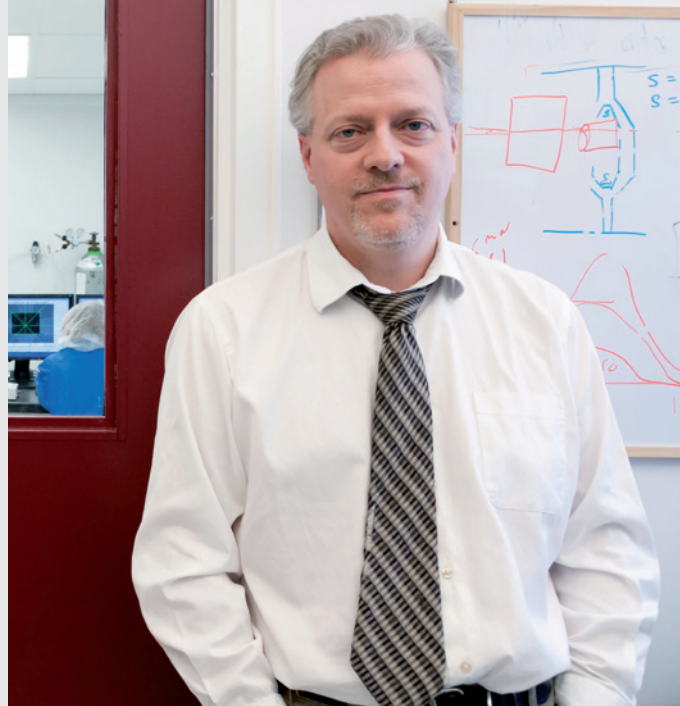


LIGHT

A *Nature* special issue
nature.com/light2015

| SQUEEZING LIGHT |

Pierre Berini



PETER THORNTON/UNIV. OTTAWA

Pierre Berini knows a bargain when he sees one; the evidence is in his lab, which is cluttered with lasers, oscillators and other components that he bought at auctions after local companies folded. The University of Ottawa physicist often buys in batches, after spotting essential items in a job lot that otherwise looks like junk. “There are lots of surprises,” he says.

Berini has a certain sympathy for the failed companies. He is a leader in plasmonics, a way of manipulating electrons with light that could be used to transmit information in super-fast computers. Months after launching a venture-backed firm, Spectalis, to market plasmonics circuitry to the communications industry in the early 2000s, he began to feel the effects of the dot-com bubble bursting. He ended up hosting an auction of his own and closing shop. Unperturbed, he plans to try again this year, launching a company to apply his technology to tiny sensors in handheld devices that detect diseases rapidly and with extreme sensitivity.

The devices use a peculiar kind of light that emerges from waves ►

he and his colleagues had not only learned how to make twisted light for themselves, but had also devised a way for it to function as an ‘optical spanner’ to trap cells and other microscopic particles, then rotate them into any position².

Turning light into a spanner is really about shaping it, Padgett says. A very simple example of shaping is a digital projector, which creates a changing image by altering a beam’s intensity pixel by pixel. A more sophisticated example is a liquid-crystal device that does nothing to the intensity of the light passing through each pixel, but instead shifts its ‘phase’ — the relative position of the wave’s peaks and troughs. In the stacked-dinner-plate analogy, the plates collectively warp and bend.

Getting to twisted light is a matter of taking that warping to extremes — so that the wavefronts form a spiral. That twist means that the beam not only exerts radiation pressure on the objects it encounters, nudging them forward, but also tries to rotate them. “It’s just like turning and pushing a door knob to open a door,” says Padgett. The optical spanner passes this momentum to microscopic objects to trap, rotate and move them. Using such devices, biologists can bump beads into cells to measure the cells’ stiffness, and engineers can create unique nanoscale materials.

Twisted light also provides a new way to encode information. The conventional approach to doing this with light is to encode each bit as a single photon spinning either clockwise or anticlockwise around its direction of motion. Quantum mechanics allows only those two possibilities, so this gives a natural way to represent the 1s and 0s of binary code.

But twisted light has an extra rotational quantity known as orbital angular momentum. This differs from intrinsic spin in the same way as Earth’s yearly motion around the Sun differs from its daily rotation on its axis. And it is

much less constrained by quantum mechanics. In theory, says Padgett, twisted light can have an infinite number of orbital angular momentum patterns, or modes, each twisted tighter than the last. “This is like having a whole alphabet with which to communicate,” he says.

A decade ago, Padgett was among the first to show that each mode can be used to encode different information³ — such as shades of grey or numbers — which allows much more data to be carried by the same optical signal than is possible with just spin encoding. Last year, a team at the University of Vienna encoded grey-scale images of Wolfgang Amadeus Mozart and other famous Austrians using 16 twisted modes, and successfully sent the images through 3 kilometres of air⁴ (see *Nature* <http://doi.org/ztt>; 2014). By using extra channels of information, such techniques could increase the data-carrying capacity of fibre-optic cables and radio waves.

Padgett has found even more imaginative ways to play with twisted light. When a beam of it illuminates a wall, for example, the spot will have a dark centre. That is because a spinning beam of light has a vortex in the middle where intensity is zero. Look closely at a spot of laser light, says Padgett, and it seems to be riddled with such dark spots, known as speckles. If you could trace these spots back through the laser beam, they would form continuous lines of zero intensity twining in three dimensions⁵. “These can be like cooked spaghetti, or you can form them into spaghetti hoops or even chain mail,” says Padgett. (He points to a poster on his wall showing what that looks like: its title is ‘Speckleghetti’.) In 2010, he and his collaborators showed how to form the lines into knots⁶. It took theorist Mark Dennis at the University of Bristol, UK,

a decade to create the complex mathematical recipe of overlapping beams needed to make an isolated, pretzel-like knot. Only with the recipe in hand could Padgett’s team use its light-shaping skills to make the abstract mathematics a physical reality.

Padgett believes that the best way for a person to succeed is for them to find something they are good at, and then to apply it everywhere. “Our team can shape light beams,” he says. “So we use shaped light in communications, microscopy, in imaging, in sensors. We always ask, how can we apply what we know to areas that others are interested in?” He is using that philosophy in his latest project: leading the Quantum Imaging Hub, a collaboration between 6 universities and

30 companies, which is one of the 4 Quantum Technology Hubs launched last November by the UK government. His group is creating infrared cameras that use a single-pixel detector rather than the millions of expensive pixels in a

conventional camera. By projecting masks of black and white squares onto an object, flickering 20,000 times a second, the team can measure how incoming intensity varies, and reconstruct a picture⁷. “It’s a convoluted, but much cheaper, way of doing the job,” says Matthew Edgar, a physicist in Padgett’s lab. With image-compression techniques and boosted computer power, the team hopes to extend the technique to video, allowing infrared cameras to spot gas leaks or see through smoke.

Back in his office and packing up to head into the Glasgow rain, Padgett reflects on what he loves about light. It is not its endless uses. Instead, he says, the beauty of light is that the more deeply you understand it, the more straightforward it gets. “If light ever surprises me, it’s not in its complexity, but that it is so simple,” he says. ■

“This is like having a whole alphabet with which to communicate.”

► of electrons propagating across a metal surface in contact with an insulator, such as air or glass. When excited with a laser, these charges, or plasmons, generate fluctuating electric and magnetic fields that flow just above the metal surface. Trapped at the interface, the waves can be funnelled into structures that confine their wavelengths to a few tens of nanometres — as little as one-tenth of the laser's wavelength. The squeezed waves travel more slowly than laser light, so can retain the same frequency.

Berini backed into studying plasmonics while looking for ways to improve normal electrical components and photo detectors in the late 1990s. Light travels much faster than electrical signals, so using it to connect silicon chips would massively speed up calculations. But light is limited by its wavelength: although electronic devices can be shrunk to a few tens of nanometres, the infrared light used in telecommunications

cannot focus to spots much smaller than a micrometre. "It's a fundamental incompatibility," says Berini. The smaller wavelengths available with plasmons looked promising, but plasmonic light does not always behave. The waves, created by the movement of electrons, decay quickly as a result of resistance in the metal, and they travel only micrometres.

Berini used tools that can craft nanoscale structures, which were becoming cheaper and more readily available, to create the first plasmonic waves that could travel for centimetres (ref. 8). His lab made whole circuits, guiding plasmons down metal strips less than 30 nanometres thick.

But allowing the waves to travel farther increases the light's wavelength. Although plasmonic waves are still smaller than conventional light waves, the compromise lessened their advantage and Berini found it tough to crack the telecommunications industry, where each component in use

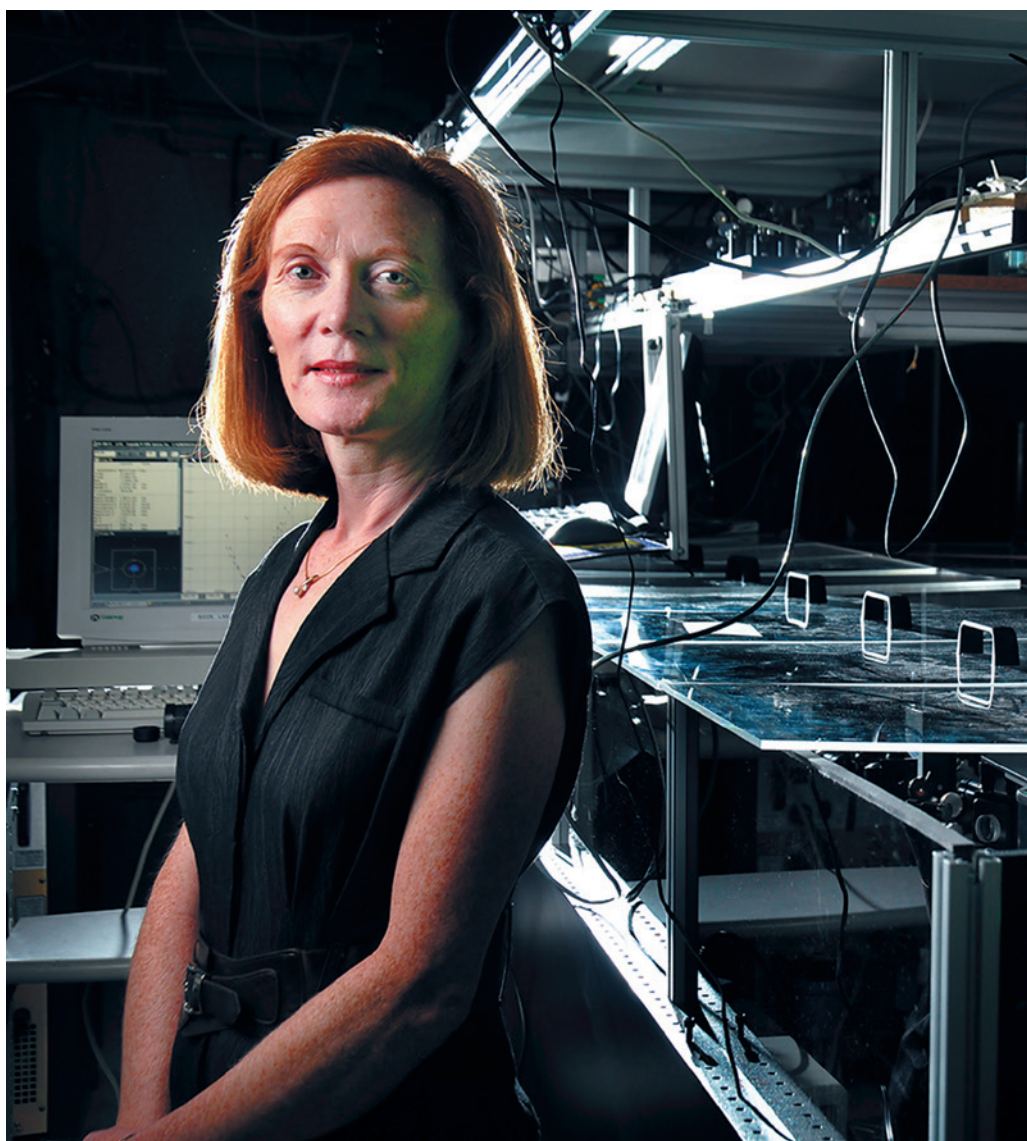
| FAST LIGHT |

Margaret Murnane

When Margaret Murnane was growing up in rural County Limerick, Ireland, in the 1960s, she had no talent for activities considered suitable for girls, such as sewing or art, and never thought of herself as being good with her hands. What she did enjoy was going on long walks with her father, and gazing at rain-drenched Ireland's multitude of rainbows — an activity that led to a lifelong fascination with light. In following that passion, she says, "it turned out I have a talent I never knew for aligning lasers. But in normal life, how would you ever know?"

Murnane's life is now that of a physicist at JILA in Boulder, Colorado, a joint institute between the University of Colorado and the US National Institute of Standards and Technology. There, with husband Henry Kapteyn, she runs a lab that is leading development of an X-ray laser that strobes in attosecond pulses, each blast lasting just one-billionth of one-billionth of a second — almost the same proportion of a second as that second is of the entire age of the Universe. Such ultrafast X-rays, which have tiny wavelengths and high energies, are often used to penetrate deep into atoms and image them at the nanometre scale. Usually, this happens at billion-dollar facilities that generate X-rays by accelerating electrons to near light speeds, such as the SLAC Linac Coherent Light Source in Menlo Park, California. By contrast, Murnane's set-up fits on a dining-room table. It allows scientists to watch the movement of electrons around atoms, probing chemical bonds or studying spins in a magnetic hard drive.

Murnane's background — a childhood spent without central heating or indoor plumbing, but with a love of knowledge and learning — lies behind much of her drive, says Kapteyn. "She worked her way up," he says. Murnane met Kapteyn as a graduate student at the University of California, Berkeley, and the two



have worked together ever since — forming a stable partnership that Murnane believes underlies their scientific success. "It helps to have someone who will challenge you hard. Those relationships are good for science, but difficult for individuals to learn," she says.

Together they tackled a problem that they first attempted in graduate school — how to

generate laser-like light beams at high energies. Rather than accelerating electrons, as huge facilities do, their strategy was to combine many visible-light photons into a handful of higher-energy X-ray photons. The process has an analogy with sound. In stringed instruments, plucking a string gently generates a single tone. "If you pluck it harder and harder,

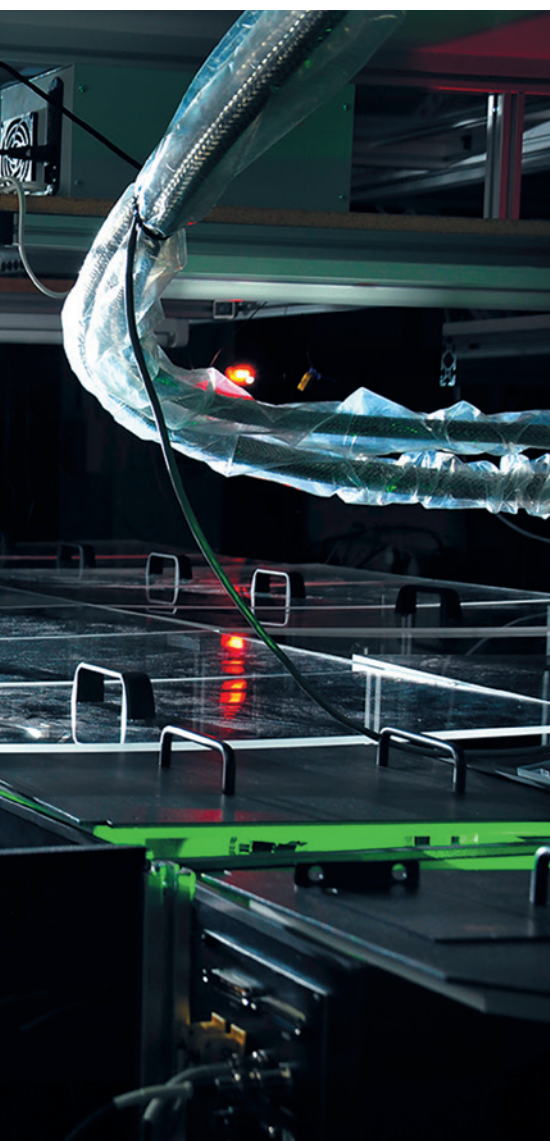
GLENN ASAKAWA/
UNIV. COLORADO

had been honed over decades. So he and others have been busy developing other techniques to deal with the short range of plasmonic light, either branching out into applications that turn the loss into an advantage, such as photodetectors, or by using nanostructures to amplify the waves. Physicists are now developing an assortment of nanoshapes — stars, rods and crescents — in a range of materials that could harness these waves for applications such as capturing solar energy, killing cancer cells and creating chip-integrated lasers, known as spasers.

Henry Schriemer, a physicist at the University of Ottawa, calls Berini a “quintessential experimentalist with a deep appreciation for the theory”. But Berini says that it is applications that turn his lab on; he attributes this entrepreneurial bent to his parents, who ran their own businesses in Timmins, the Ontario mining and logging community where he grew up.

Today, Berini is recycling the efforts made in long-range circuits to make a detector for dengue fever. The device, a handheld biosensor developed last year with researchers at the University of Malaya in Kuala Lumpur, sends plasmon waves down a chip scattered with dengue virus particles. A blood sample is placed on the chip; if the donor has the infection, the sample will contain antibodies that bind to the virus, disrupting the wave and producing a signal⁹. Berini says that the sensors could speed up diagnosis, which normally involves sending samples away to a lab.

A new company is now in the works to commercialize a range of similar biosensors. But Berini believes that the application is just one of many that squeezed light will have in the future. “With plasmonics, there is a lot of new physics to be uncovered,” he says. All of which means that some of the random equipment that litters the lab might find a new use. ■



higher harmonics emerge,” says Murnane, each at larger integer multiples of the original frequency. When ultrashort-pulse lasers were developed in the 1990s, Murnane and Kapteyn realized that they might be able to use them to ‘pluck’ an electron violently — accelerating it away from and back towards an atom of helium — and thereby generate harmonics in the form

of higher-energy photons. The team succeeded in making bright ultraviolet beams¹⁰, but it was more difficult to increase the energy while keeping the beam laser-like, with the waves emerging in synchrony.

Murnane often says that she picked physics “because it was the hardest subject” at university — an attitude that stood her in good stead with this challenge, which took 15 years to solve. The solution was to engage in what she calls “a very different way of thinking”, and start not with visible-light lasers, but with longer-wavelength infrared lasers. The photons had much less energy than before. But they resonated much more strongly with the electrons in the helium atoms — in effect, giving the string a much stronger pluck — which allowed the team to combine more than 5,000 laser photons into a single X-ray photon. Theorists believed that the technique would be too inefficient to make usable beams. But by carefully tuning the helium gas so that the laser and X-rays travelled at the same speed, Murnane’s team predicted, then proved, that the X-rays would emerge in step, as a bright beam¹¹. “What was amazing was not just that they got the X-rays, but that they got plenty of them,” says Mikhail Ivanov, a physicist at the Max Born Institute for Nonlinear Optics and Short Pulse Spectroscopy in Berlin.

Murnane and Kapteyn have now made ultrafast lasers that produce X-rays of up to 1,000 electronvolts in energy, and in attosecond pulses. Although these devices do not reach the energies or brightness attained at the big free-electron laser facilities, they come close. And, at US\$1 million, they are around one-thousandth of the price. The lab at JILA has eight such lasers, and discoveries in the nano-world are starting to trickle in. Murnane both builds and uses the lasers — processing the X-ray scatter patterns to capture images of charge and spin flows within materials. One counter-intuitive finding is that nanometre-sized heat sources cool quicker when packed closer together¹². Murnane, together with collaborators including Kapteyn and Tenio Popmintchev at JILA

and Andrius Baltuska at the Vienna University of Technology, is still working on refining the desktop set-up to make it even faster, more energetic and smaller. That would allow them to probe even quicker processes, deeper within materials and with higher resolution. “We’re pretty optimistic we can do it,” says Murnane.

After visible lasers were invented in 1960, they underwent rapid development; the same revolution is now happening for tabletop X-ray sources. Other labs around the world have

developed similar approaches, says Olga Smirnova, a theorist at the Max Born Institute. But what makes the JILA technique stand out is the ability to produce such high-frequency light, with such efficiency. And then there is Murnane herself, says Smirnova: “She is really able

“She is really able to push the envelope of what is possible, year after year.”

to push the envelope of what is possible, year after year.”

Murnane insists that they have not reached the limit yet — that higher-energy X-rays and even faster, zeptosecond (10^{-21} s) pulses may be possible. “A misconception in science sometimes is that lasers are now an old technology, and there’s nothing new to learn,” she says. “That’s so far from the truth. ■

Elizabeth Gibney is a reporter for Nature in London.

- Allen, L., Beijersbergen, M. W., Spreeuw, R. J. C. & Woerdman, J. P. *Phys. Rev. A* **45**, 8185 (1992).
- Simpson, N. B., Dholakia, K., Allen, L. & Padgett, M. J. *Opt. Lett.* **22**, 52–54 (1997).
- Gibson, G. *et al. Opt. Express* **12**, 5448–5456 (2004).
- Krenn, M. *et al. New J. Phys.* **16**, 113028 (2014).
- O’Holleran, K., Dennis, M. R., Flossmann, F. & Padgett, M. J. *Phys. Rev. Lett.* **100**, 053902 (2008).
- Dennis, M. R., King, R. P., Jack, B., O’Holleran, K. & Padgett, M. J. *Nature Phys.* **6**, 118–121 (2010).
- Sun, B. *et al. Science* **340**, 844–847 (2013).
- Charbonneau, R., Berini, P., Berolo, E. & Lisicka-Shrzek, E. *Opt. Lett.* **25**, 844–846 (2000).
- Wong, W. R., Krupin, O., Sekaran, S. D., Mahamud Adikan, F. R. & Berini, P. *Anal. Chem.* **86**, 1735–1743 (2014).
- Durfee, C. G. III, Backus, S., Murnane, M. M. & Kapteyn, H. C. *Opt. Lett.* **22**, 1565–1567 (1997).
- Popmintchev, T. *et al. Science* **336**, 1287–1291 (2012).
- Hoogetboom-Pot, K. *et al. Springer Proc. Phys.* **162**, 341–344 (2015).



SUPER VISION

USING TECHNIQUES ADAPTED FROM ASTRONOMY, PHYSICISTS ARE FINDING WAYS TO SEE THROUGH OPAQUE MATERIALS SUCH AS LIVING TISSUE.

by Zeeya Merali

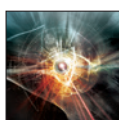
It seemed too good to be true, says Allard Mosk. It was 2007, and he was working with Ivo Vellekoop, a student in his group at the University of Twente in Enschede, the Netherlands, to shine a beam of visible light through a 'solid wall' — a glass slide covered with white paint — and then focus it on the other side. They did not have a particular application in mind. "I really just wanted to try this because it had never been done before," Mosk says. And in truth, the two researchers did not expect to pick up much more than a faint blur.

But as it turned out, their very first attempt¹ produced a sharp pinprick of light a hundred

times brighter than they had hoped for. "This just doesn't happen on the first day of your experiment," exclaims Mosk. "We thought we'd made a mistake and there must be a hole in our slide letting the light through!"

But there was no hole. Instead, their experiment became the first of two independent studies^{1,2} that were carried out that year pioneering ways to see through opaque barriers. So far it is still a laboratory exercise.

But progress has been rapid. Researchers have now managed to obtain good-quality images through thin tissues such as mouse ears³, and are working on ways to go deeper. And if they can meet the still-daunting challenges, such as



LIGHT

A *Nature* special issue
nature.com/light2015

ILLUSTRATION: VIKTOR KOEN

dealing with tissues that move or stretch, potential applications abound. Visible-light images obtained from deep within the body might eliminate the need for intrusive biopsies, for example. Or laser light could be focused to treat aneurysms in the brain or target inoperable tumours without the need for surgery.

"Just ten years ago, we couldn't imagine high-resolution imaging down to even 1 centimetre in the body with optical light, but now that has now become a reality," says Lihong Wang, a biomedical engineer at Washington University in St. Louis, Missouri. "Call me crazy, but I believe that we will eventually be doing whole-body imaging with optical light."

RICH SOURCE

It is already possible to peer inside the body with X-rays and ultrasound. But the images produced by such tools are crude compared with those that should be possible with visible light. Partly this is because visible-light images tend to have higher resolution, says Wang. But it is also because optical wavelengths interact strongly with organic molecules, so the reflected light is packed with information about biochemical changes, cellular anomalies and glucose and oxygen levels in the blood.

However, those interactions also make visible light prone to scattering and absorption. Absorption will scupper any imaging attempt: the information the photons pick up is lost as they are absorbed into the material. Scattering, however, preserves a ray of hope. Many materials, such as skin, white paint or fog, are 'opaque' only because photons passing through them ricochet until they are thoroughly scrambled. But they are not lost — so in principle, the scrambling can be reversed.

Astronomers have already solved a version of this scattering problem using a technology called adaptive optics, which allows them to undo the distortions imposed on images of stars, planets and galaxies by the scattering of light in the atmosphere (see *Nature* **517**, 430–432; 2015). The basic idea is to collect light from a bright reference star and use an algorithm to calculate how the atmosphere has smeared and blurred its point-like image. The algorithm then controls a special 'deformable' mirror that cancels out the atmospheric distortions, turns the guide-star image into a point, and at the same time brings other distant objects into sharp focus.

Unfortunately, this technique is tough to use in the body. Targets deep inside biological tissues do not shine the way that stars do — they have to be illuminated from the outside — and the scatterers are much more densely packed than those that scatter light in the atmosphere. "You'd need the equivalent of a deformable mirror with billions of moving parts to compensate for the scattering caused by an egg shell," says Ori Katz, an optical physicist at the Langevin Institute in Paris. That is why Mosk and Vellekoop were not too hopeful of success when they started. Still, the pair took heart from the advance of technology. "Until recently it had been preposterous to think you could control a million pixels, but, by 2007, every smartphone could do it," says Mosk.

They therefore made use of a 'spatial light modulator': a device similar to an LCD smartphone display that can control the transmission of different parts of a laser beam by delaying one part relative to another. They fired their laser through the modulator towards the painted glass slide, placed a detector beyond the slide and used a computer to monitor how much light the detector picked up. The computer then added and subtracted delays at each pixel of the modulator, going through a process of trial and error to see what changes minimized the scattering of the laser light as it passed through the slide. In effect, it was trying to give the incoming light a distortion that the opaque barrier would exactly cancel out. Mosk and Vellekoop ran the algorithm for more than an hour, and when it was done they had a result that beat all their expectations: a focus that was a thousand times more intense than the background signal¹.

"The Mosk experiment was an eye-opener," says Katz. "It changed the paradigm of what could be done with optical light."

Soon after his success, Mosk learned of similar work being done by bioengineer Changhui Yang and his team at the California Institute of Technology in Pasadena.

These researchers had used a different technique to focus scattered optical light, and a different opaque substance: a thin slice of chicken breast². But they, too, were surprised by how easy it was to do. "I had thought 'we'll spend six months on this, and when it doesn't work, we'll chalk it up as a learning experience,'" says Yang. "But actually it wasn't that hard."

Soon after the two papers were published, the field exploded as other physicists rushed to join in. One of them was optical physicist Jacopo Bertolotti, who came to work with Mosk in 2010. Bertolotti, now at the University of Exeter, UK, says that he was drawn both by the "beauty of the experiment" and by the potential it offered for medical imaging. But he could see that that goal was still a long way off.

"CALL ME CRAZY, BUT I BELIEVE THAT WE WILL EVENTUALLY BE DOING WHOLE-BODY IMAGING WITH OPTICAL LIGHT."

The first issue that Bertolotti faced was that Mosk's original set-up required a camera to be placed behind the opaque surface. That is a problem for medical applications because placing a camera under the skin would involve surgery, which would be invasive, dangerous and rarely worth the risk. In 2012, however, Bertolotti, Mosk and their colleagues devised a way to put both the laser light source and the detector in front of the surface⁴.

Their target was a fluorescent Greek letter π just 50 micrometres across hidden behind a thin opaque screen. As such, the target was roughly the same size as a cell and analogous with medical techniques that involved injecting fluorescent dyes into living tissue to aid in imaging. When the laser was switched on, the photons would bounce their way through the screen and produce a diffuse illumination of the fluorescent π . The light reflected from the letter would then make its way back through the screen and produce a blurry speckled pattern on the other side. It was like trying to see the symbol through a shower curtain.

Yet the shape of the letter was still encoded in the scattered light. To retrieve that shape, the team recorded the speckle pattern, moved the laser to shine at a different angle, then recorded the new speckle pattern⁴. By repeating this many times and comparing the patterns point by point, a computer could work out how the patterns were correlated — and from that, work backwards to reconstruct the hidden letter π .

That was progress, says Bertolotti, but it still was not good enough. "It only works if the object to be imaged is on the other side of the scattering medium," he says. For many medical applications, such as seeing inside the brain, or within a blood vessel, the target is buried within tissue.

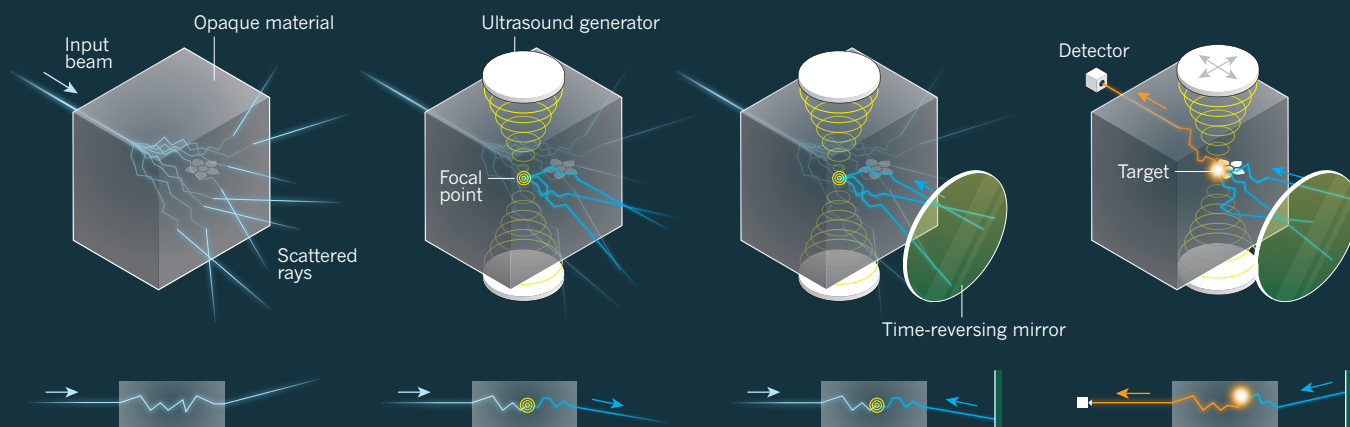
INSIDE OUT

The challenge of imaging inside the scattering medium has been taken up by a number of groups, including Yang's and Wang's. In 2013, for instance, Yang's team achieved this feat with unprecedented resolution by picking out a fluorescent bead just one micrometre across sandwiched between two artificial opaque layers⁵.

Yang, together with biologist Benjamin Judkewitz and the rest of his team did this by illuminating the medium and letting the light bounce its way through to the other side, then reflecting it back with a 'time-reversing' mirror, which effectively forces every light ray to exactly retrace its steps. Time-reversing all the rays would simply undo all the scattering, however. So instead, the team focused an ultrasound beam — which is not easily scattered — at one point in the medium, knowing that any optical light that happened to pass through that point would

LIGHT AND SOUND

One way to see inside opaque materials is to combine ultrasound with a 'time-reversing' mirror system, which forces every light ray to exactly retrace its steps.



SCRAMBLED LIGHT

In many opaque materials — including living tissue, white paint and fog — light is not actually absorbed. It simply bounces around in the material until it is too scrambled to form an image.

ULTRASOUND FOCUS

A beam of ultrasound (yellow rings) is focused at some point within the material. Any light that happens to bounce through this point will undergo a slight shift in its frequency (blue rays).

TIME REVERSAL

A time-reversing mirror sends back only the frequency-shifted light. When the light retraces its steps, it passes through the ultrasound focal point and adds its energy to the light coming through on the first pass.

SCANNING THE MEDIUM

The focal point is scanned through the material. When it passes over targets labelled with a fluorescent dye, the structures emit a detectable glow — and researchers can build up a map of the interior.

undergo a slight shift in frequency. Then on the far side, the researchers set up the time-reversing mirror tuned so that it would send back only the light that had experienced that frequency shift. The result was a thin, time-reversed beam that would automatically pass back through the focus and add its energy to the light from the first pass. This turned the ultrasound focus into a spot of comparatively high radiation intensity — “a torch inside the wall”, says Judkewitz, who is now at the Charité University Hospital in Berlin. Better still, the ultrasound focus could be moved around within the medium. And when it passed over the bead, the bead fluoresced (See ‘Light and sound’).

However, the technique was still a long way from seeing into deep layers of tissue, which pose another, much tougher challenge: they tend to move constantly as a result of blood flow and breathing. “We are still not so close to medical applications because these techniques tend to work only if the scattering medium is perfectly frozen in time,” says Mathias Fink, a physicist at Langevin who pioneered a version of the time-reversal technique in the 1990s that used ultrasound alone⁶. Most groups have reduced the timing from Mosk’s original hour or so to just tens of seconds, says Katz, and that is fine for imaging a bead or a letter π , but not for imaging a tumour in the body.

But last year, a team led by Sylvain Gigan, a physicist at the Kastler Brossel Laboratory in Paris, and including Katz and Fink, demonstrated a way to reconstruct the image of the hidden object in just one camera shot⁷. “It’s a bit like magic when you see the algorithm converge on the final image,” Gigan says.

Wang agrees that speed is of the essence. “Everything is in motion and we only have a millisecond-scale window to make an image,” he says. In a paper published in January³, Wang and his team managed to get the speed down to 5.6 milliseconds, “which is fast enough for selected *in vivo* imaging”, he says. Furthermore, their target was made from ink-stained gelatin and sandwiched between the ear of an anaesthetized mouse and a ground-glass diffuser. Getting success with a live mouse is impressive, says Bertolotti — although he points out that “moving from a mouse ear, which is relatively thin, to imaging human skin and flesh will still take a lot of extra work”.

As of today, Bertolotti adds, there is still no imaging approach that

stands out above the rest. Each has its advantages and disadvantages. “Rather than developing one technique that’s good for everything, I think we’ll develop a suite of techniques that could one day all be combined into the same piece of apparatus,” he says. “I don’t know how quickly that might happen, but this is a young and fast-moving community, so it could be within a few years.”

The techniques now being pioneered by bioengineers and physicists for medicine could also be put to a range of other purposes. Mosk, for example, believes that these methods could be a tool for art restoration. “Most painters build up works in several layers, and the layers below can influence the chemical and physical ageing of the painting, so it’s of some significance that you know what is in there if you want to preserve it,” he says. Methods that in effect unscattered light could also help the telecommunications industry to unscramble the noise in optical fibres that is caused by scattered light.

Another obvious customer is the military, says Fink, who thinks that the technology could be used to allow soldiers to see through a portable shield — either a physical screen or a fogging spray — that obscures them from their enemy’s view. “It’s not the same as being invisible, but it would allow you to see others while not being seen,” he says.

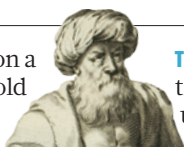
Almost all the scientists in this young field get excited when they start dreaming of applications. But Gigan, for one, is keen to keep the applications above board. “When we tell people what we do, someone always asks if we’ll create a phone app to let people look through shower curtains,” he says. “This is something that could be done with our technique — but we don’t intend to do it.” ■

Zeeya Merali is a freelance writer in London.

1. Vellekoop, I. M. & Mosk, A. P. *Phys. Rev. Lett.* **101**, 120601 (2008).
2. Yaqoob, Z., Psaltis, D., Feld, M. S. & Yang, C. *Nature Photon.* **2**, 110–115 (2008).
3. Liu, Y. *et al. Nature Commun.* **6**, 5904 (2015).
4. Bertolotti, J. *et al. Nature* **491**, 232–234 (2012).
5. Judkewitz, B., Wang Y. M., Horstmeyer, R., Mathy, A. & Yang, C. *Nature Photon.* **7**, 300–305 (2013).
6. Cassereau, D. & Fink, M. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **39**, 579–592 (1992).
7. Katz, O., Heidmann, P., Fink, M. & Gigan, S. *Nature Photon.* **8**, 784–790 (2014).

COMMENT

LIGHT Reflections on a thousand-year-old optics text **p.164**



THEATRE Tom Stoppard tackles the science of consciousness, using fraud and finance **p.166**

CONSERVATION Fishing lobbyists re-open debate over Brazil's Red List **p.167**

OBITUARY Hubert Markl, architect of German science system, remembered **p.168**

RICHARD DREW/AP



A software glitch at trading firm Knight Capital caused losses of US\$440 million in a single day in 2012.

Trading at the speed of light

To minimize risks, we must learn more about how financial markets operate at ever faster rates, urges **Mark Buchanan**.

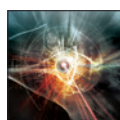
Financial traders are in a race to make transactions ever faster. In today's high-tech exchanges, firms can execute more than 100,000 trades in a second for a single customer. This summer, London and New York's financial centres will become able to communicate 2.6 milliseconds (about 10%) faster after the opening of a transatlantic fibre-optic line dubbed the Hibernia Express, costing US\$300 million. As technology advances, trading speed is increasingly limited only by fundamental physics, and the ultimate barrier — the speed of light.

Through glass optical fibres, information travels at two-thirds of the speed of light in

a vacuum (300,000 kilometres per second). To go faster, data must travel through the air. The corridors between Chicago and New York and New Jersey, and between London and Frankfurt, are bristling with efficient microwave and millimetre-wave links. An even more efficient network of lasers — based on military technology for in-flight signalling between aeroplanes — has been installed to link the New York and New

Jersey as well as the London and Frankfurt financial exchanges¹.

Next up may be hollow-core fibre cables, through which light would travel in a tiny air gap at light speed. Trading firms speculate about a fleet of balloons or uncrewed solar-powered drones carrying signal repeaters to support a network of links across the oceans. In a decade or so, firms may even communicate using neutrinos, which travel at the speed of light and can go through obstacles, including Earth. It all spells big profits for high-tech trading firms, which now account for around 50% of equity trading in the United States and in Europe. ▶



LIGHT

A *Nature* special issue
nature.com/light2015



Laser units on rooftops connect New Jersey's Nasdaq data centre with the New York Stock Exchange.

► But some firms claim that uneven access to extreme speed erodes trading fairness. And system-wide failures occur when algorithms interact in unforeseen ways — such as in the ‘flash crash’ of 6 May 2010, when the Dow Jones Industrial Average fell by the largest daily amount ever within minutes (see ‘Flash crash’). No one knows when a similar event might spill over into global markets.

Avoiding these risks will require intensive research on how markets work — as complex ecologies of interacting algorithms — and how countermeasures could avert disasters.

GETTING AHEAD

High-frequency trading relies on fast computers, algorithms for deciding what and when to buy or sell, and live feeds of financial data from exchanges. Every microsecond of advantage counts. Faster data links between exchanges minimize the time it takes to make a trade; firms fight over whose computer can be placed closest; traders jockey to sit closer to the pipe. It all costs money — renting fast links costs around \$10,000 per month.

Communications technology is a limiting factor. Fibre-optic cables carry the most data, but do not give the speed required. The fastest links carry information over a geodesic arc — the shortest path on Earth's surface between two points. So line-of-sight microwaves are a better option; millimetre waves and lasers are better yet, because they have higher data densities.

Open-air communications systems are prone to weather disruption. Anova Technologies, a network provider for trading firms headquartered in Chicago, Illinois, has augmented its New York laser network with

millimetre waves to overcome rain, fog and snow. Adaptive alignment mechanisms keep the links working even if winds make towers twist by up to 3°. But microwaves and lasers cannot be used over long distances without repeaters. They attenuate quickly in the atmosphere and do not curve around Earth.

Some economists question the worth of such investments. Joseph Stiglitz, a Nobel laureate in economics, is among those who argue that rapid trading is socially useless². High-frequency firms quickly cancel about 95% of the orders they make³. Worse, speed may impede proper market function. The traditional purpose of financial markets is to pool diverse information from many people to channel investment resources. That requires trading based on insight, depth of study and patience — all foreign to the high-frequency algorithm-based system².

GOOD, BAD AND UGLY

Fast trading has pros and cons. First, it gives markets ‘liquidity’ — it makes it easier for investors to find trading partners at reasonable prices. Liquid markets benefit trade in the same way that free-flowing traffic helps transport. Such markets tend to have low ‘spreads’ — the difference between the prices at which one can buy or sell a stock, which reflects the fee that dealers demand and thus transaction costs for investors. As high-frequency trading has grown over the past decade, spreads in many markets have fallen, making trading cheaper⁴.

Even so, the liquidity that computer trading creates is fleeting, and it can fail when markets get unruly. Wildly fluctuating prices mean bigger risks for traders who earn a living by ‘market making’ — standing ready to buy or sell stocks at any moment and earning

a profit from the spread. The algorithms they use to trade profitably make more errors and are programmed to get out of the market altogether when markets get too volatile. The problem is exacerbated by the similarity of the algorithms used by many high-frequency trading firms — they all bail out at the same time. That is what happened in the 2010 flash crash. (Of course, this problem happens with human traders too, who flee markets when they get too scary.)

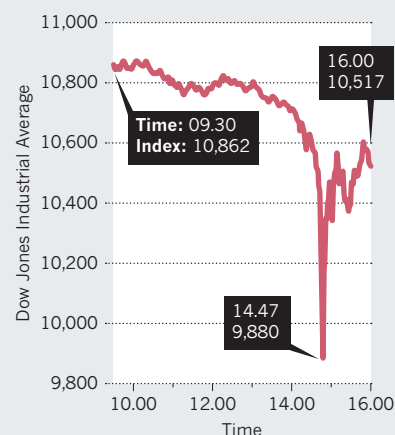
Another good thing about high-frequency trading is that it helps to synchronize prices across markets⁵. It takes time to digest information, draw out implications and align prices. If prices in sugar or high-fructose corn syrup rose, stocks in Coca Cola would fall quickly; those of less well-known soft-drink companies would take longer. High-frequency trade speeds up that process. In 2000, it took minutes on average for a price change in one security to flow to others. Now it takes less than ten seconds. Not everyone likes this: fast synchronization wipes out profit opportunities for firms that make money by knowing about the momentary price imbalances.

MARKET DYNAMICS

Some high-frequency firms exploit an anachronism in the structure of markets. By US law, each regulated exchange must feed its best available prices for a stock, sale and purchase, to a central facility, which uses that information to establish a public National Best Bid and Offer (NBBO). But exchanges also sell faster proprietary data feeds that firms can use to predict the NBBO in advance, gaining an edge over anyone using the public information alone. Hence, high-frequency firms can move in ahead of slower traders. This tends to further synchronize prices. Big investors such as mutual funds and pension

FLASH CRASH

On 6 May 2010, the market value of the Dow Jones Industrial Average index fell by 9%, but recovered in minutes. High-speed trading algorithms were in part to blame.



SOURCE: DOW JONES

funds, which act on real-world insight and information with a long-term view, are among those which lose out, although they also benefit from the lower spreads created by high-frequency traders.

In the United States, some large trading firms have set up private trading spaces to eliminate the timing edge for high-frequency traders. For example, the alternative trading system IEX, launched in 2013, aims to stop exploitation of the NBBO. It has introduced a trading 'speed bump' — an automatic delay of 350 microseconds — which makes it impossible for traders to benefit from the faster feeds. IEX has already attracted about 1% of stock-trading volume in the United States. Firms in other countries may follow suit.

With computer codes carrying out trades with real-world consequences at a rate beyond that at which humans can intervene, the impacts of coding errors and digital glitches can spiral quickly. In 2012, a flaw in the algorithms of one of the largest US high-frequency trading firms, Knight Capital, caused losses of \$440 million in 45 minutes as its system bought at higher prices than it sold.

Sudden spikes or 'fractures' in the prices of stocks are increasingly common. Tens of thousands of times in the past few years, stock values have changed by 1% in less than 0.04 of a second⁶. The flash crash of 2010 happened at around 2.45 p.m. New York time, and markets recovered in about 15 minutes. Had it struck just before closing time in New York, the shock would have affected markets worldwide and recovery would have taken longer. Some investors speculate about a 'splash crash', in which a massive spike in one market disrupts or freezes trade in foreign exchange, futures, commodities, bonds and other assets, potentially triggering a global economic crisis.

Some researchers⁶ suggest that the spikes reflect a fundamental transformation of market dynamics, linked to the necessity for firms to use simple algorithms to maximize running speed.

SYSTEMIC RISKS

The nature of financial markets today is vastly different from that in the past. Rather than reflecting the collective decisions of people, they belie the behaviour of complex webs of technologies and their interactions with humans. The potential for global problems is increasing as high-frequency trading has moved into international markets for futures and other assets⁷. No industry — including energy and food, insurance and banking — is immune from disruption.

In future, when airborne laser networks span the oceans, things may get even stranger. The location at which traders get the earliest possible information from two exchanges lies at their mid-point — between

FAST TRADING HOTSPOTS

The speed of light is the ultimate limit to how rapidly trades can be made between financial centres (●) — it would take signals travelling at this speed 67 milliseconds to travel halfway around the Earth. The midpoints between exchanges (•) are the best places to site high-frequency trading computers because they access information from both simultaneously and with the minimum delay.



Chicago and London, this is in the middle of the Atlantic Ocean. At such a site, traders could exploit a technique called 'relativistic arbitrage'⁸ to profit from momentary imbalances in prices in Chicago and London.

To explain: special relativity says that nothing can travel faster than the speed of light, c . Hence, a trader standing a distance D away from an exchange can find out what happened there, in the best circumstance, at a time $T = D/c$ after it happened. Between major trading centres around the globe, such delays can be from a few to tens of milliseconds. If a trader stands halfway between the two exchanges, he or she will receive information from both after the same interval, $T = D/c$. Anywhere else, the distance to at least one of the exchanges would be greater and information would take longer to get there.

In other words, within a few years it may become profitable to station a ship or

"The impacts of coding errors and digital glitches can spiral quickly."

other trading platform near halfway points between pairs of financial centres worldwide (see 'Fast-trading hotspots'). That said, the profits earned by high-frequency firms have

fallen in recent years, suggesting that most of the easy opportunities for money-making have already been taken. If in ten years the wheels of the global financial system really will be greased by firms signalling from New York to Melbourne at Einstein's speed limit, research and policy-making should focus on two questions. First, how to avoid the biggest things that can go wrong; and second, how to make markets work as well as they can to serve society.

The first challenge requires more research

into the dynamics of markets that are run by algorithms rather than investors. Computer scientists, mathematicians and economists need to work together to understand what drives flash crashes and how changes in market structures might avoid them. What 'circuit breakers', so to speak, might keep events from running out of control?

Second, researchers and policy-makers need to assess how to regulate markets to make them serve the purpose of boosting real economic investment. Algorithmic trading has been given wide latitude for the past two decades, under the assumption that firms making a profit must be helping the market. Finance research⁹ suggests that there may be an optimal speed for trading that today's markets have already far surpassed. ■

Mark Buchanan is a science writer based in the United Kingdom. His latest book is *Forecast: What Physics, Meteorology and the Natural Sciences Can Teach Us About Economics*.

e-mail: buchanan.mark@gmail.com

1. Patterson, S. 'High-Speed Stock Traders Turn to Laser Beams' *The Wall Street Journal* (12 February 2014).
2. Stiglitz, J. E. 'Tapping the Brakes: Are Less Active Markets Safer and Better for the Economy?' Paper given at 2014 Financial Markets Conference (2014); available at <http://go.nature.com/ushstd>.
3. Salmon, F. 'The problem with high frequency trading' *Reuters* (2012); available at <http://go.nature.com/ejnucs>.
4. Hendershott, T., Jones, C. M. & Menkveld, A. J. *J. Finance* **66**, 1–33 (2011).
5. Gerig, A. Preprint at <http://arxiv.org/abs/1211.1919> (2012).
6. Johnson, N. et al. Preprint at <http://arxiv.org/abs/1202.1448> (2012).
7. Cliff, D. & Northrop, L. *The Global Financial Markets: An Ultra-Large-Scale Systems Perspective* (UK Government Office for Science, 2010).
8. Wissner-Gross, A. D. & Freer, C. E. *Phys. Rev. E* **82**, 056104 (2010).
9. Fricke, D. & Gerig, A. Preprint available at <http://dx.doi.org/10.2139/ssrn.2363114> (2015).

IN RETROSPECT

Book of Optics

Jim Al-Khalili revisits Ibn al-Haytham's hugely influential study on its millennium.

The greatest physicist of the medieval era led a life as remarkable as his discoveries were prodigious, spending a decade in prison and at one point possibly feigning mental illness to get out of a tight spot. Abu Ali al-Hassan ibn al-Haytham (Latinized to Alhazen) was born in Basra, now in southern Iraq, in AD 965. His greatest and most famous work, the seven-volume *Book of Optics* (*Kitab al-Manazir*) hugely influenced thinking across disciplines from the theory of visual perception to the nature of perspective in medieval art, in both the East and the West, for more than 600 years. Many later European scholars and fellow polymaths, from Robert Grosseteste and Leonardo da Vinci to Galileo Galilei, René Descartes, Johannes Kepler and Isaac Newton, were in his debt. Indeed, the influence of Ibn al-Haytham's *Optics* ranks alongside that of Newton's work of the same title, published 700 years later.

Interest in optics began in antiquity. The Babylonians, Egyptians and Assyrians all used polished quartz lenses. The basic principles of geometric optics were laid down by Plato and Euclid. They included ideas such as the propagation of light in straight lines, and simple laws of reflection from plain mirrors. The earliest serious contribution from the Islamic world came from ninth-century Arab scholar Ya'qub ibn Ishaq al-Kindi.

As a young man, Ibn al-Haytham received an excellent education and was widely noted as a mathematical and scientific prodigy. Frustrated by his administrative duties working in a government post in the vast Islamic Empire — which at the time stretched from India to Spain — he was sacked owing to real or, as some speculate, faked mental illness.

Sometime during the first decade of the new millennium, he proposed an ambitious project to dam the Nile. He was invited to Egypt by the Fatimid caliph al-Hakim bi'amr Allah. However, on seeing the scale of the task, Ibn al-Haytham quickly realized that it was beyond him. He was promptly imprisoned in Cairo for wasting the caliph's time.

Far from cowering him, the decade of imprisonment granted Ibn al-Haytham the seclusion to think and write, particularly on optics. After his release around the year 1020, he began working at a prolific rate, carrying out a series of



Ibn al-Haytham's work long pre-dated Newton's.

famous experiments on the nature of light. For example, using a camera obscura, he proved that light travels in straight lines; he also mathematized the fields of catoptrics (reflection of light by mirrors) and dioptrics (refraction of light through lenses). This huge body of experiment and theory culminated in his *Book of Optics*.

This treatise can be regarded as a science textbook. In it, Ibn al-Haytham gives detailed descriptions of his experiments, such as exploring how light rays are reflected off plain and curved surfaces. He includes the apparatus he used, the way he set it up, the measurements and his results. He then uses these observations to justify his theories, which he develops with geometrical models. He even urges others to repeat his experiments to verify his conclusions. Many historians of science consider Ibn al-Haytham to be the first true proponent of the modern scientific method.

The work can be roughly divided into Books I, II and III, devoted to the theory of

vision and the associated physiology of the eye and the psychology of perception; and Books IV to VII, covering traditional physical optics. The work's most celebrated contribution to science is its explanation of vision.

At that time, scholars' understanding of the phenomenon was a mess. The Greeks had several theories. In the fifth century BC, Empedocles had argued that a special light shone out of the eye until it hit an object, thereby making it visible. This became known as the emission theory of vision. It was 'refined' by Plato, who explained that you also need external light to see. Plato's student Aristotle suggested that rather than the eye emitting light, objects would 'perturb' the air between them and the eye, triggering sight. Other philosophers around this time, including Epicurus, attempted a form of 'intromission theory' of vision (light entering the eye from outside), but it was Plato's theory that was given a mathematical basis by Euclid, who described light rays emerging in a cone from the eye. Several centuries later, Ptolemy expanded on this idea.

Early Islamic scholars such as al-Kindi and Hunayn ibn Ishaq favoured a combined emission-intromission theory. They posited that the eye sends out light to the observed object, which then reflects the light back into the eye.

It took the genius of Ibn al-Haytham to finally resolve the issue. He argued that if we see because rays of light are emitted from the eye onto an object (Plato and Euclid's 'sight rays'), then either the object sends back a signal to the eye or it does not. If it does not, how can the eye perceive what its rays have fallen on? Light must be coming back to the eye, and this is how we see. But if so, what use is there for the original rays emitted by the eye? The light could come directly from the object if it is luminous or, if it is not, could be reflected from the object after being emitted by another source. Rays from the eye, decided Ibn al-Haytham, are an unnecessary complication.

He also went further than anyone before in trying to understand the underlying physics of refraction. He argued that the speed of light was finite and varied in different media, and he used the idea of resolving the path



LIGHT

A *Nature* special issue
nature.com/light2015

of a light ray into its vertical and horizontal components of velocities. He carried out all his work geometrically, and introduced many new ideas, such as the study of how the atmosphere refracts light from celestial bodies.

Later Islamic scholars, including the thirteenth-century Persians Qutb al-Din al-Shirazi and Kamal al-Din al-Farisi, extended the *Optics*. Al-Farisi, who wrote *The Revision of the Optics* (*Tanqih al-Manazir*), used geometry to arrive at the first correct mathematical explanation of the rainbow (at the same time as, but independently of, the German scholar Theodor of Freiberg).

The *Book of Optics* was first translated into Latin in the late twelfth or early thirteenth century, as *De Aspectibus*. The English philosopher and empiricist Roger Bacon then wrote a summary of it, as did his Polish contemporary Witelo. It was soon being cited across Europe. Among the many ideas taken up by Ibn al-Haytham's Latin-reading disciples was that pure light was not visible, and that its job was simply to allow us to see

“Many historians of science consider Ibn al-Haytham to be the first true proponent of the modern scientific method.”

colour. Even Kepler, who studied Ibn al-Haytham's work, thought this; it took Newton to describe light as itself being made up of different colours. (Other erroneous ideas in *Optics* include a repetition of

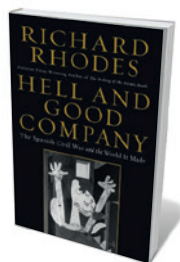
Ptolemy's mistaken law of refraction, and an incorrect understanding of reflection as a more intense form of refraction.)

Ibn al-Haytham's work decisively influenced the theory of perspective that flowered in Renaissance European science and art. *De Aspectibus* was translated into Italian in the fourteenth century, making it accessible to practitioners such as the Florentine art theorist and architect Leon Battista Alberti, author of the 1435 treatise *On Painting* (*Della pittura*), the sculptor Lorenzo Ghiberti and the geometer-artist Piero della Francesca. They harnessed Ibn al-Haytham's discussions on perspective to help to create the illusion of three-dimensional depth on canvas and in friezes. These revolutionary artists strove to understand both the objective world and the visual system that determined its appearance.

Today, as we use laser beams to manipulate atoms, stimulate neurons with light or convey information in entangled photons, it is worth recalling that the foundations of this field were laid down around 1,000 years ago by Ibn al-Haytham. ■

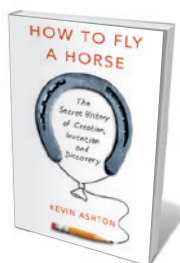
Jim Al-Khalili is a professor of physics at the University of Surrey in Guildford, UK, and the author of *Pathfinders: the Golden Age of Arabic Science*.
e-mail: j.al-khalili@surrey.ac.uk

Books in brief



Hell and Good Company: The Spanish Civil War and the World it Made
Richard Rhodes SIMON AND SCHUSTER (2015)

His 1986 *The Making of the Atomic Bomb* (Simon and Schuster) is a towering chronicle of modernity. Now historian Richard Rhodes examines the “little world war”, Spain's hellish 1936–39 civil conflagration (see *Nature* **494**, 34; 2013). As he shows, it was a testing ground for medical and technological advances — in blood transfusion on one hand, and on the other in airborne warfare, which led to the bombing of Guernica immortalized in Pablo Picasso's great painting. Luminaries drawn to the war, he shows, ranged from geneticist J. B. S. Haldane to writer George Orwell.



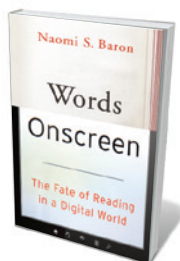
How to Fly a Horse: The Secret History of Creation, Invention, and Discovery
Kevin Ashton DOUBLEDAY (2015)

This study of creativity by Kevin Ashton — the technical pioneer behind the ‘Internet of Things’ — is a testament to Thomas Edison's definition of genius (1% inspiration, 99% perspiration). Science, Ashton argues, is less a thing of “eureka shrieks” than of hard work, small steps and understanding of adversity. His case studies compel, from Réunion slave Edmond Albius's 1841 breakthrough in vanilla pollination to surgeon Judah Folkman's “series of repetitive failures” that led to the discovery of angiogenesis, now key to cancer treatments.



Future Arctic: Field Notes from a World on the Edge
Edward Struzik ISLAND (2015)

In September 2014, dwindling sea ice forced some 35,000 walrus onto the Alaskan coast — just one ecological event in a multitude besetting the ‘climate-changed’ Arctic. Journalist and explorer Edward Struzik cogently analyses the environmental and policy challenges, drawing on research into past extinctions and present disruptions such as tar-sand exploitation, military territorialism and tundra fires. As he ticks off the costs to indigenous peoples, ocean biodiversity, caribou habitat and more, the case for an Arctic treaty and serious conservation efforts becomes ever clearer.



Words Onscreen: The Fate of Reading in a Digital World
Naomi S. Baron OXFORD UNIVERSITY PRESS (2015)

For every digital devotee clutching an e-reader, there is an old-school bibliophile brandishing a physical book. But which works best for reading comprehension? In this thoughtful study, linguist Naomi Baron investigates each platform in the light of recent research, and surveys US, Japanese and German reading habits. E-readers, she finds, democratize access and offer easy storage, but can also discourage tackling more involved texts or rereading, and encourage “power browsing” rather than perusal. She recommends allowing room for both options — letting “form follow function”.



What Nature Does For Britain
Tony Juniper PROFILE (2015)

Part research round-up, part manifesto, this treatise on Britain's ‘natural capital’ is a model of pragmatism. As environmentalist Tony Juniper shows, UK ecosystems were valued at £1.6 trillion (US\$2.4 trillion) in 2011 by the Office for National Statistics. Yet poor practices such as overfishing and soil degradation are breaking nature's bank. Juniper offers smart policy action points for switching to sustainability, and ingenious case studies — from ‘woodland system’ farming to reintroduced beavers that help with riverine flood control. [Barbara Kiser](#)

Correspondence

Greek politics stall research reforms

The ongoing damage to Greek scientific research is not solely due to austerity measures (*Nature* **517**, 127–128; 2015). In my experience as a member of Greece's National Council for Research and Technology from 2010 to 2014, political manipulation and institutional weakness are also contributors.

To address the dire problem of underfunded research, in 2011 the council introduced an open, competitive grant scheme (called *Aristeia*, or excellence) based on the European Research Council model. It ran for two rounds, during which we had to battle against other governmental forces to maintain its European Union (EU) funding. The scheme now seems to have been abandoned.

The council developed a multi-annual plan that year for research and development (R&D) to bring Greece closer to EU expenditure targets by 2020. This was stalled and diverted by government. We pressed for the creation of a high-level government committee to oversee R&D, and for a research agency similar to the US National Science Foundation. That plan was also lost, diluted by the research law you mention.

The council's experience reflects the wider problems of Greece's government: how it seeks and receives expert advice, the public status of this process and the near-impossibility of rational, stable long-term planning. The shallow and short-term strictures of the 'troika' — the three organizations that act for Greece's creditors — make matters worse.

Kevin Featherstone *London School of Economics & Political Science, UK.*
k.featherstone@lse.ac.uk

Leave Brazil's Red List alone

Brazil's government has agreed to review its updated Red List of the country's threatened marine

species. This review represents a victory for lobbyists in the fishing industry. It is not based on new biological information.

The list that was issued in December 2014 by Brazil's environment ministry (through decrees 444 and 445) was the culmination of a six-year process involving 1,300 national and international scientists, overseen by the International Union for Conservation of Nature. It restricts or bans the capture of several commercially valuable fish, such as groupers and sharks.

The fishermen's unions last month questioned the criteria for inclusion and persuaded Helder Barbalho, minister of fisheries and aquaculture, and environment minister Izabella Teixeira to review the list.

A repeal of decree 445 or an amendment of the Annex I list, which regulates the capture of 409 fish species and 66 aquatic invertebrates, would be a serious setback for conservation and for the sustainable management of fisheries in Brazil.

Alexander C. Lees *Museu Paraense Emílio Goeldi, Belém, Pará, Brazil.*
alexanderlees@btopenworld.com

EU research plan may widen gaps

The success of the European Union (EU) Research and Innovation programme depends on achieving critical mass among member states and optimizing each state's research contribution (see also M. Żylicz *Nature* **517**, 438; 2015). This could be difficult, given the wide variation in each state's willingness to participate and in their investment in research.

Research excellence and competitiveness remain concentrated in just a few geographical areas, despite efforts by the EU to promote homogeneity. It is those regions that make the advances in research and technology, fuelling

the imbalance (see K. Schwab (ed.) *The Global Competitiveness Report 2013–2014* World Economic Forum, 2013).

The EU plan to align national research programmes could make matters worse. Closer cooperation between researchers and between states will help to secure research sponsorship and collaboration with scientists outside Europe. But these advantages are more likely to be enjoyed by high-performing countries, further widening the gap from the others. The proposed alignment will also have to struggle with extra bureaucracy and delays (M. Cuijpers *et al. Res. Policy* **40**, 565–575; 2011).

Pier Francesco Moretti *National Research Council of Italy, Brussels, Belgium.*
pierfrancesco.moretti@cnr.it

Biodiversity: include freshwater species

The omission of freshwater species from your biodiversity assessment (*Nature* **516**, 158–161; 2014) reflects a more general bias towards terrestrial conservation, borne of insufficient knowledge about freshwater ecosystems. The Red List of the International Union for Conservation of Nature, for example, is dominated by freshwater fish species whose population status is unknown.

Integrative conservation measures are particularly important in places where people depend on freshwater resources for subsistence, and where human activities are rapidly changing rivers, lakes and their surrounding landscapes. The highly biodiverse Amazon River basin is an example. In parts of Africa, diminishing supplies of freshwater fish have led to the overexploitation of terrestrial animals (J. S. Brashares *et al. Science* **306**, 1180–1183; 2004).

We need more data for freshwater ecosystems to inform

conservation strategies and to integrate them with terrestrial habitats.

Sebastian Heilpern *University of Chicago, Illinois, USA.*
sheilpern@uchicago.edu

Biodiversity: sharks and rays in peril too

Your status report on fauna biodiversity (*Nature* **516**, 158–161; 2014) overlooks a group that is causing serious concern among conservationists — sharks, rays and chimaeras. These are particularly vulnerable to fishing and by-catch, in part because they mature late and produce few young.

An estimated 24% of this group, known as chondrichthyan fish, are threatened with extinction under the Red List criteria of the International Union for Conservation of Nature. This exceeds the percentage for birds and is comparable to that for mammals. There are insufficient data to determine status in 47% of chondrichthyan fish, and models predict that many of these could also be under threat, given their similar life history and morphology to the listed chondrichthyans.

Extinction of ocean fish is hard to verify. There is as yet no documented global extinction of a chondrichthyan, but many populations are locally or regionally extinct (such as sawfishes (Pristidae family); see N. K. Dulvy *et al. Aquat. Conserv.* <http://doi.org/zkc>; 2014). Some critically endangered species, including the Pondicherry shark (*Carcharhinus hemiodon*) in the Indo-West Pacific, have not been recorded in decades and may already be extinct.

Peter M. Kyne *Charles Darwin University, Darwin, Australia.*
Nicholas J. Bax *CSIRO, Australia; and University of Tasmania, Hobart, Australia.*
Nicholas K. Dulvy *Simon Fraser University, Burnaby, Canada.*
peter.kyne@cdu.edu.au

Hubert Markl

(1938–2015)

Biologist who steered German research organizations through reunification.

Hubert Markl had an extraordinary impact on research in Germany before, and crucially during, the turbulent process of reunification. An evolutionary biologist and behavioural scientist, he was also a writer, public intellectual and policy-maker. His stints as president of the German Research Foundation (1986–91), the Berlin-Brandenburg Academy of Sciences and Humanities (1993–95), and the Max Planck Society (1996–2002) shaped the entire German and European research systems.

Markl died on 8 January, aged 76. He was born in Regensburg, southern Germany, in 1938. Although he had an early interest in the humanities, Markl studied biology, chemistry and geography at Ludwig Maximilian University in Munich. His teachers included luminaries such as the behavioural scientists Martin Lindauer, Konrad Lorenz and Karl von Frisch and the zoologist Hansjochem Autrum. He got his doctorate in zoology aged 24.

During the early 1960s, Markl held several research posts in the United States: at Harvard University in Cambridge, Massachusetts, Rockefeller University in New York and the Tropical Research Station of the New York Zoological Society (where colleagues called him Jim). He returned to Germany, to the Goethe University Frankfurt. In 1967, he submitted a thesis on the communication behaviour of social insects to acquire his lecturing qualification.

In 1968, he became professor and director of the zoological institute at Darmstadt University of Technology. Markl recalled this appointment as the most crucial and successful of his life. It gave him the freedom to pursue research interests from evolutionary biology and behavioural ecology to sensory physiology and conservation. In 1974, Markl moved to the University of Konstanz, founded eight years before to revive the Humboldtian ideal of research-based teaching. He became one of the leading figures of 'Little Harvard on Lake Constance'.

That year, Markl was also elected senator of the German Research Foundation, the nation's main public funding agency for basic research. After a six-year stretch as vice-president, he became its youngest ever president in 1986. Of his many achievements there, three stand out: his implementation of long-term grants; the introduction of a structured programme for doctoral training and research;



and the opening up of funding opportunities for East German researchers well before unification was agreed on in the autumn of 1990.

Next, Markl became deeply involved in unifying Germany's two higher-education and research systems that had headed in different directions after the Second World War. In West Germany, teaching and research were combined in a federal system where each state had a lot of independence. The East had adopted the Soviet model of universities tooled mainly for teaching and specialist institutes focused on research. In 1993, Markl's task as founding president of the Berlin-Brandenburg Academy of Sciences and Humanities was to attract the best researchers from the former East and West to become active members of a unified academy. He forged joint working groups through which the best minds came to trust each other.

Markl faced much bigger challenges when he took the helm of Germany's Max Planck Society in 1996. He was the first and so far only president recruited from outside the organization. The society had planned 18 new institutes in eastern Germany. But owing to the government's severe underestimation of the costs of unification, the organization did not get the funds it needed. Hard decisions were required.

It was clear to Markl that savings had to be made at existing institutes in the west, and more resources transferred to the new institutes in the east. This controversial policy

quickly earned him a tough reputation, particularly when he decided to close underperforming and outmoded departments, as well as entire Max Planck institutes, such as the one for history in Göttingen and for cell biology at Ladenburg, near Heidelberg. The closures were resisted by the affected state governments. With his sharp intellect and his talent for communication, Markl prevailed and rejuvenated the Max Planck Society.

During his term, 153 new directors out of the society's 266 were appointed. As a result of a root-and-branch evaluation of the society, Markl improved the institutes' links with neighbouring universities, such as Göttingen, Munich and Heidelberg. In 2000, he started the International Max Planck Research Schools programme. The scheme has attracted several thousand young scholars from abroad to study in Germany and continues to build bridges across institutional boundaries. Many Max Planck directors have become closely involved in training doctoral students as well as in the teaching and research activities at the respective German partner universities.

Markl spoke truth to power on topics including genetic engineering, cloning and stem-cell research. He was also outspoken against xenophobia and in favour of intercultural learning and the right to medically assisted suicide.

In 1997, he initiated an independent study of the history of the Kaiser Wilhelm Society (the predecessor of the Max Planck Society from 1911 to 1946) during the Third Reich. In 2001, as a result of this research, he publicly acknowledged the guilt of its members participating in the expulsion of Jewish colleagues and other Nazi atrocities, and apologized to survivors at a commemoration ceremony.

"Responsibility does not rest with science as such," Markl repeatedly told his students, "it is always the individual scientist." He will long be remembered as a visionary, a brilliant intellectual and courageous leader. Without him, scholarship and science in Germany and beyond would not be what they are today. ■

*Wilhelm Krull is secretary-general of the Volkswagen Foundation in Hanover, Germany. He frequently collaborated with Hubert Markl over 30 years.
e-mail: krull@volkswagenstiftung.de*

WOLFGANG FLISER/MPG

FORUM Cosmology

The oldest cosmic light

The cosmic microwave background is a faint glow of light left over from the Big Bang. It fills the entire sky and records the Universe's early history. Two independent experts outline what we know about this ancient light, both theoretically and observationally.

Life history of the photon

DAVID SPERGEL

The oldest photons in the Universe are the cosmic background radiation (CBR). These photons are fossils formed in the first hours after the Big Bang. In the first 300,000 years of cosmic history, protons, ions and electrons formed a dense plasma. The CBR photons scattered off these electrons just like light scatters off a dense fog. During this early epoch, any form of energy injection would have produced further photons and distorted the energy spectrum of the CBR. The COBE satellite's measurement of this spectrum, which showed no detectable deviation from the thermal (black-body) form¹ and was one of the achievements that led to the 2006 Nobel Prize in Physics, constrains the Universe's early history and is one of the pillars of the Big Bang theory.

During the next 100,000 years, electrons and protons combined to make neutral hydrogen. Because hydrogen is transparent to the CBR photons, these photons could start to propagate freely. They have travelled for about

13.8 billion years, the approximate age of the Universe. By the time they reach our detectors, they have redshifted to microwave wavelengths, and so we observe them as the cosmic microwave background (CMB) radiation. The temperature and polarization patterns of the CMB bear the signatures of the photons' last interactions with electrons; the polarization describes the direction of the electric fields that the photons carry.

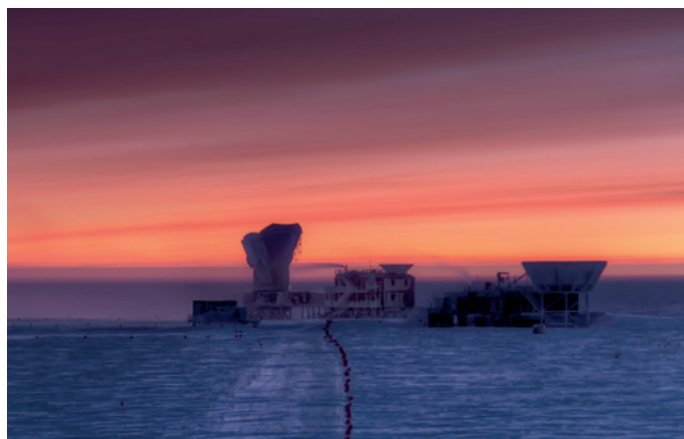
According to the most popular cosmological model, the early Universe underwent a period of exponential expansion called inflation. During this inflationary expansion, the volume of the Universe grew by more than 180-exponential-fold, and tiny fluctuations in the light's quantum field were amplified into density fluctuations on the scales of billions of light years. These density fluctuations generated sound waves that propagated through the early Universe. The sound waves produced a distinctive pattern of ripples that have been seen not only in the measurements of temperature fluctuations in the CMB by the Wilkinson Microwave Anisotropy Probe (WMAP)^{2,3}, the Planck⁴ satellite and ground-based CMB experiments, but also in the Sloan Digital Sky Survey⁵ measurements of the large-scale distribution of galaxies.

A remarkably simple model, a Universe filled with dark matter, atoms and dark energy, and seeded by the inflationary fluctuations, beautifully fits these observations. These precise measurements determine the basic parameters of the cosmos: its age, density, shape and composition. The statistical properties of the fluctuations provide an extra test of inflation.

The sound waves also produce a distinctive pattern of fluctuations in the polarization of the CMB. Polarization fluctuations can be divided into two types: 'E-mode' polarization — patterns that are symmetric under mirror reflection and could be created by the variations in density produced by inflation, and swirly 'B-mode' fluctuations. The WMAP and Planck satellites have seen the predicted E-mode patterns, another observational triumph for the theory of inflation.

Despite the remarkable observational successes of this theory, there are several profound unsolved theoretical problems associated with the inflationary model. Alternative models exist. Our next-best observational test of the model is to detect primordial gravitational waves — ripples in the fabric of space-time generated during the Universe's early rapid expansion. Brian Keating's companion article

STEFFEN RICHTER, HARVARD UNIV.



BRIAN KEATING



Figure 1 | Searching for the origins of the Universe. Several telescopes around the world, including three at the South Pole (the Keck Array, BICEP2 and the South Pole Telescope; all three shown left) and two in the Atacama Desert in Chile (right; the Atacama Cosmology Telescope towards the top of the picture and the POLARBEAR telescope in the foreground), are dedicated to observing the cosmic microwave background (CMB) — relic radiation from the Big Bang. Each telescope tackles CMB observations in a different way. The image on the right also shows the future site of the Simons Array, currently under construction.

describes the distinctive B-mode signature of the gravitational waves.

David Spergel is in the Department of Astrophysical Sciences, Princeton University, Princeton, New Jersey 08544, USA.
e-mail: dns@astro.princeton.edu

Massless messengers

BRIAN KEATING

The CMB is the most perfect black body known in nature¹, much better than any laboratory black-body oven, whose walls are almost perfect photon absorbers. These ancient CMB photons are also gravimeters. Using their intensity and polarization properties, we can measure the gravitational field of the last scattering surface — the fictitious shell formed by the Universe's first hydrogen atoms. Like the walls of an oven, the last scattering surface is so absorptive that it is the end of the line-of-sight, the farthest we can look back, at least using photons. However, the last scattering surface is also a gravitational-wave detector: a thin 'film' of matter on which primordial gravitational waves can be exposed, allowing us to peer back to earlier epochs when these waves themselves were produced. If inflation produced gravitational waves, then the waves will have imprinted a unique pattern of B-mode polarization on the CMB^{6–8}. As described in David Spergel's companion article, if the B-mode polarization pattern proves to be of primordial origin, it will be strong evidence that inflation occurred. This is the goal of nearly a dozen CMB polarimeters that are either planned or currently plying Southern Hemisphere skies (Fig. 1).

CMB polarimeters are astonishingly precise. Current experimental sensitivities are at the level of tens of nanokelvin, unimaginable just a decade ago when the hunt for B modes began⁹. This is due to a 'Moore's-law-like' growth in the number of detectors. But these are no ordinary smartphone pixels. CMB polarimeters use superconducting bolometers, thermal sensors cooled below 300 millikelvin. Two bolometers, one per polarization state of the CMB, are coupled to reflecting (mirror-based) or refracting (lens-based) telescopes. Although details vary by instrument, polarimeters exploit the differential nature of the signal — what matters is the difference in microwave power between the CMB's two polarization states. To detect this difference, experimentalists cleverly use

the twofold modulation of polarized signals for each single physical rotation about the optical axis of the polarimeter. (Try it with polarized sunglasses on a sunny day at sunset. Looking at the zenith, spin around once: the sky brightness modulates twice.)

But experimental challenges run deeper than raw sensitivity. Systematic effects masquerading as primordial gravitational waves need be only a few parts per billion of the ambient 300 K background to swamp observations. Despite this extreme susceptibility, astrophysical foregrounds such as dust emission from the Milky Way now seem to present the most formidable challenge.

Polarization data obtained with the BICEP2 telescope in Antarctica were initially interpreted as evidence for inflation¹⁰, but this conclusion was questioned¹¹ and recently reinterpreted. A joint analysis using Planck observations at a frequency of 353 gigahertz and data from the Keck Array and BICEP2 at 150 GHz has shown that BICEP2's original B-mode data were likely to be not purely primordial, nor caused by systematics. Rather, they were potentially dominated by thermal emission from dust grains in the Milky Way aligned by Galactic magnetic fields¹². The game is still afoot — only experiments observing the cleanest celestial regions, with sensitivity in multiple frequency bands, can unambiguously detect the inflationary signal.

The prospect of detecting gravitational waves from the inflationary epoch, 10^{−36} seconds after the Big Bang, is exhilarating. And yet, more-exotic physics may lurk undiscovered in the CMB's polarization. As CMB photons traverse the cosmos, they trace the properties of matter (dark and luminous)

and the curvature of space-time itself. CMB photons are lensed; that is, their trajectories are bent by gravitational fields produced by dark matter. Such CMB polarization lensing perhaps offers the best hope of measuring the mass of the neutrino¹³, the only elementary particle whose mass is unknown.

Ancient photons are illuminating fundamental cosmic mysteries. CMB photons shed light on subjects that were once only the purview of particle colliders: elementary particle masses and ultra-high energy fields. The next 100 years promise to be equally exhilarating, thanks to the indomitable, massless messenger of the cosmos — the photon. ■

Brian Keating is at the Center for Astrophysics & Space Sciences, Department of Physics, University of California, San Diego, La Jolla, California 92093-0424, USA.
e-mail: bkeating@ucsd.edu

1. Fixsen, D. J., Hinshaw, G., Bennett, C. L. & Mather, J. C. *Astrophys. J.* **486**, 623–628 (1997).
2. Spergel, D. N. *et al. Astrophys. J. Suppl.* **148**, 175–194 (2003).
3. Bennett, C. L. *et al. Astrophys. J. Suppl.* **208**, 20 (2013).
4. Planck Collaboration. *Astron. Astrophys.* **571**, A1 (2014).
5. Aubourg, E. *et al.* Preprint at <http://arxiv.org/abs/1411.1074> (2014).
6. Polnarev, A. G. *Sov. Astron.* **29**, 607–613 (1985).
7. Seljak, U. & Zaldarriaga, M. *Phys. Rev. Lett.* **78**, 2054–2057 (1997).
8. Kamionkowski, M., Kosowsky, A. & Stebbins, A. *Phys. Rev. Lett.* **78**, 2058–2061 (1997).
9. Keating, B. G. *et al. Proc. SPIE* **4843**, 284–295 (2003).
10. BICEP2 Collaboration. *Phys. Rev. Lett.* **112**, 241101 (2014).
11. Flauger, R., Hill, J. C. & Spergel, D. N. *J. Cosmol. Astropart. Phys.* JCAP08(2014)039 (2014).
12. BICEP2/Keck, Planck Collaborations. Preprint available at <http://arxiv.org/abs/1502.00612> (2015).
13. Abazajian, K. N. *et al. J. Astropart. Phys.* <http://dx.doi.org/10.1016/j.astropartphys.2014.05.014> (2014).

CELL BIOLOGY

Death drags down the neighbourhood

An analysis of dying cells reveals that they play an active part in modifying tissue shape by pulling on neighbouring cells. This induces neighbouring cells to contract at their apices, which results in tissue folding. SEE LETTER P.245

CLAUDIA G. VASQUEZ & ADAM C. MARTIN

A normal and essential part of tissue development, maintenance and repair is apoptotic cell death¹, in which cells shrink and fragment into membrane-bound structures called apoptotic bodies that are engulfed by phagocytic cells. Apoptosis has a key role in sculpting tissue morphology, removing cells from between forming digits and eliminating vestigial structures. However, apoptosis is often regarded as a passive elimination

of unwanted cells. On page 245 of this issue, Monier *et al.*² report a surprising finding — apoptosis can trigger contractions that fold tissues. Rather than being inert, apoptotic cells actively affect their surroundings, causing lasting changes in tissue form and structure.

Epithelial tissues line the body's cavities, creating boundaries between different extracellular environments. Accordingly, epithelial cells are polarized, with one side of the cell dubbed apical and the other basal. In epithelial tissues, apoptosis in one cell induces



LIGHT

A Nature special issue
nature.com/light2015

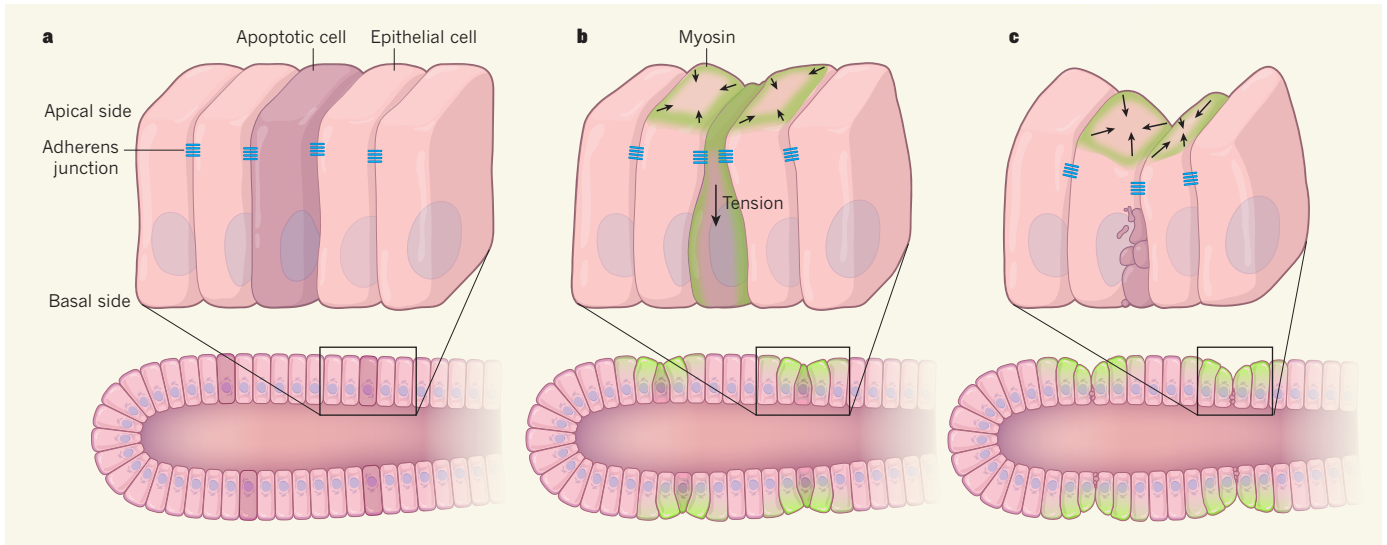


Figure 1 | Apoptosis and fold formation. Apoptotic cell death in an epithelial-cell sheet causes folding in the forming leg joints of fruit-fly larvae. **a**, Cells in this epithelial sheet are attached to their neighbours at structures called adherens junctions. The junctions remain intact even as cells undergo apoptosis. **b**, Monier *et al.*² report that apoptotic cells in the epithelial sheet

shrink and accumulate myosin protein along their apical–basal axis as they do so. Neighbouring cells also accumulate myosin at their apical surfaces, causing elevated tension around the apoptotic cell. **c**, As the dying cell fragments, neighbouring cells apically constrict and form a fold in the tissue.

apical contractility in neighbouring cells. The result is a purse-string-like contraction that pushes the dying cell from the epithelium³. Furthermore, epithelial cells remain adhered to their neighbours at specific structures called adherens junctions while they undergo apoptosis, inducing neighbouring cells to stretch and elongate towards the dying cell⁴. Although apoptosis is also known to be required to generate tissue-wide shape changes⁵, the link between this shape change and contractility is poorly understood.

Contractile forces are generated through the actin and myosin proteins that make up a cell's cytoskeleton. Actin filaments assemble into meshes and bundles that underlie cell membranes, whereas myosin is a motor protein that forms mini-filaments that both connect and 'walk' along actin filaments. Myosin mini-filaments can contract the actin network to generate cellular tension, and this contractility can be increased and transmitted between cells through adherens junctions in response to signalling molecules, promoting dramatic changes in tissue shape.

Such cellular tension is essential for the generation and maintenance of specific tissue architectures. One well-characterized example of how cell contractility and the resulting change in cell shape can influence tissue architecture occurs in the fruit fly *Drosophila melanogaster*. During early embryonic development, a signalling pathway activates the actin–myosin cytoskeleton, constricting the apices of a specific set of epithelial cells. This folds the epithelial sheet and results in the movement of muscle-cell precursors inside the embryo^{6,7}. By contrast, the cytoskeletons of cells in a mature, static epithelium constantly tug on neighbouring cells. This state of tension

is required to maintain an ordered hexagonal cell array in epithelial tissues⁸. Thus, a fundamental property of epithelial tissues is that cells continually exert pulling forces on each other.

The leg joints of fruit flies form in the third larval stage of the insects' development, when folds in epithelial tissue cause successive rounds of tissue subdivision. This process involves apoptosis in cells at the centre of the fold⁵ and contraction of the apical side of the surrounding tissue, which then folds inward towards the basal side. Monier and colleagues set out to determine the link between apoptosis and fold formation in these leg joints. Using live imaging, they studied a fluorescent version of myosin in cells undergoing apoptosis, and report that the protein accumulates along the apical–basal axis of the apoptotic cell. They also observed that the apoptotic cell pulls on its neighbours as it contracts and shrinks into the epithelium. Consequently, apical myosin levels in neighbouring cells increase, which causes these cells to constrict their apices and form a fold (Fig. 1).

How does apoptosis cause neighbouring cells to constrict? The authors show that inhibiting myosin function in the apoptotic cell suppresses the accumulation of myosin at the apical edges of surrounding cells, and also suppresses fold formation². This suggests that the mechanical pulling force from the dying cell could trigger myosin activity in its neighbours.

Studies^{9,10} have demonstrated that applying mechanical forces to tissues can elevate myosin activity. But the mechanism through which a pulling force exerted by an apoptotic cell could recruit apical myosin in the surrounding tissue remains unknown. One possibility is that the myosin motor itself responds to tension, with tension increasing the length of time that myosin remains bound to actin

filaments¹¹. Alternatively, tension might affect signalling factors that regulate activity of the enzyme Rho kinase (as has been proposed for folding embryonic epithelia¹⁰), because myosin dynamics and spatial organization are regulated by a balance between myosin phosphorylation by Rho kinase and dephosphorylation by phosphatase enzymes¹². It is also possible that contractile activity in the apoptotic cell compromises the integrity of the cell's membrane, leading to the release of a chemical signal, such as ATP, that induces contraction in neighbours¹³. One experiment that would directly test whether physical transmission of tensile force is required to elevate myosin would be to disrupt adherens-junction proteins, which are required to mechanically couple epithelial cells during apoptosis⁴.

How common is the phenomenon of apoptosis triggering or contributing to morphogenetic movements? Apoptotic cells contribute to generating the tension that drives epithelial-sheet movement in a fruit-fly tissue called the amnioserosa¹⁴. Monier and co-workers' study demonstrates that apoptosis — and possibly the resulting pulling force — induces myosin accumulation and fold formation in three other epithelial tissues. Thus, apoptotic cells may have a more active role in shaping tissues than was previously believed.

Tissue folding is most often thought to be triggered by transcription factors or secreted signalling molecules, or both. This study demonstrates that in some tissues there is in fact a relay effect, in which induction of one type of cell behaviour can trigger changes in the surrounding tissue. The idea that mechanical signalling can trigger a propagation of contractile activity was originally proposed in some of the first mechanical models of tissue

folding more than 30 years ago¹⁵. It is surprising that this trigger could represent a dying cell's final tug. ■

Claudia G. Vasquez and Adam C. Martin are in the Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02142, USA. e-mail: acmartin@mit.edu

1. Baehrecke, E. H. *Nature Rev. Mol. Cell Biol.* **3**,

- 779–787 (2002).
 2. Monier, B. *et al. Nature* **518**, 245–248 (2015).
 3. Rosenblatt, J., Raff, M. C. & Cramer, L. P. *Curr. Biol.* **11**, 1847–1857 (2001).
 4. Lubkov, V. & Bar-Sagi, D. *Curr. Biol.* **24**, 868–874 (2014).
 5. Manjón, C., Sánchez-Herrero, E. & Suzanne, M. *Nature Cell Biol.* **9**, 57–63 (2006).
 6. Dawes-Hoang, R. E. *et al. Development* **132**, 4165–4178 (2005).
 7. Manning, A. J., Peters, K. A., Peifer, M. & Rogers, S. L. *Sci. Signal.* **6**, ra98 (2013).
 8. Farhadifar, R., Röper, J.-C., Aigouy, B., Eaton, S. & Jülicher, F. *Curr. Biol.* **17**, 2095–2104 (2007).
 9. Fernandez-Gonzalez, R., de Matos Simoes, S.,

- Röper, J., Eaton, S. & Zallen, J. A. *Dev. Cell* **17**, 736–743 (2009).
 10. Pouille, P.-A., Ahmadi, P., Brunet, A.-C. & Farge, E. *Sci. Signal.* **2**, ra16 (2009).
 11. Ren, Y. *et al. Curr. Biol.* **19**, 1421–1428 (2009).
 12. Vasquez, C. G., Tworoger, M. & Martin, A. C. J. *Cell Biol.* **206**, 435–450 (2014).
 13. Herrgen, L., Voss, O. P. & Akerman, C. J. *Dev. Cell* **31**, 599–613 (2014).
 14. Toyama, Y., Peralta, X. G., Wells, A. R., Kiehart, D. P. & Edwards, G. S. *Science* **321**, 1683–1686 (2008).
 15. Odell, G. M., Oster, G., Alberch, P. & Burnside, B. *Dev. Biol.* **85**, 446–462 (1981).

This article was published online on 21 January 2015.

STAR FORMATION

Sibling rivalry begins at birth

High-resolution astronomical observations of a nearby molecular gas cloud have revealed a quadruplet of stars in the act of formation. The system is arguably the youngest multiple star system detected so far. SEE LETTER P.213

KAITLIN M. KRATTER

Roughly 150 years after the invention of the telescope, astronomers deciphered the mysteries of double stars. Using probability arguments, natural philosopher John Michell suggested in 1767 that pairs of stars in the sky were in gravitational tangos, not chance alignments. The early discovery of such binary stars reflects their numbers: at least half of stars like the Sun are found in multiple systems^{1,2}. And yet the origins of this all-too-normal population are mysterious. On page 213 of this issue, Pineda *et al.*³ report the discovery of a quadruple star system still in the ‘womb’ (Fig. 1). With these data, we are closing in on understanding binary-star conception.

Why worry about binary stars versus single stars at all? It turns out that binary stars have an outsized impact on many astrophysical phenomena. First, binaries allow us to detect stellar-mass black holes, which are the remnants of massive stars and can be seen only because of their gravitational influence on a normal stellar companion. Second, they generate the type Ia supernovae explosions that are used as ‘standard candles’ to measure cosmic distances. Third, we now know that planets can form in binaries, notably in the habitable ‘Goldilocks’ temperature zone, where the planet is at the right distance from the host star to retain at least some liquid water. Finally, binaries may also be a source of gravitational waves. For all of these research areas, characterizing the binary population is crucial, and proper accounting must start at birth.

Young stars have more bound companions than their older counterparts^{4,5}. This observation indicates that binary stars form together

at birth, rather than uniting after they grow up. Pineda and colleagues’ infant quadruple system reinforces this idea. Multiple and single stars alike form in cold, filamentary molecular gas clouds. Within filaments, we find protostars forming in dense regions because of the triumph of gravitational forces over thermal, turbulent and magnetic pressures. The relative importance of turbulence and magnetic fields in star formation is an open question⁶. Theoretical models that require turbulence to set the properties of star-forming regions predict that sometimes there is not just one dominant dense region in a filament, but two or three, or

even four^{6–9}. Frequently, these regions turn into a bound protostellar system. As they continue to accrete gas and dust from their surroundings on their way to becoming fully fledged stars, sometimes sibling rivalry plays a part in deciding which star gets to be the biggest¹⁰. The siblings either work out their differences and settle into organized periodic orbits, or, if necessary, eject one party to keep the peace.

Pineda and colleagues have observed arguably the youngest multiple system so far using the Karl G. Jansky Very Large Array in New Mexico and the James Clerk Maxwell Telescope in Hawaii. They discovered four distinct dense gas condensations in a nearby molecular cloud, only one of which hosts a protostar. All four objects are incredibly young in astronomical terms, probably less than 10⁵ years old (based on their densities and temperatures).

Detecting systems at birth is extremely difficult on multiple counts. First, the star-formation timescale is only a few hundred thousand years, which is short compared to stellar lifetimes of billions of years. Thus, the probability of ‘catching’ star systems in the act of formation is small. Second, these siblings can be seen through their molecular womb only with observations that have high

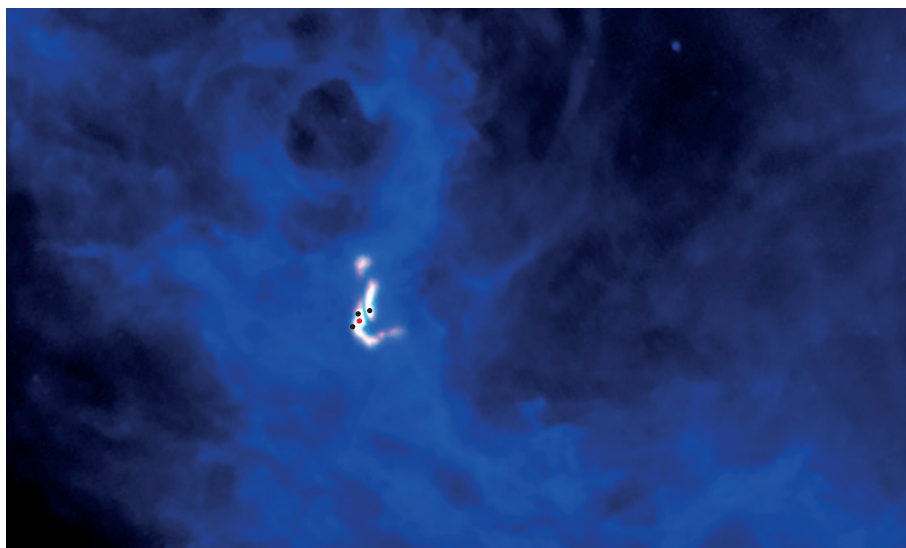


Figure 1 | A young quadruplet. Pineda *et al.*³ have discovered four distinct gas condensations in a clumpy, filamentary gas cloud (white) surrounded by dust (blue). The locations of the condensations in this image are marked with black and red dots. The four condensations are destined to form a bound multiple star system, and one of them (red dot) has already ‘turned on’ as a protostar.

sensitivity and high spatial resolution to separate one body from the next. Given the relative rarity of quadruple star systems at older ages, one might think this discovery improbable, or lucky. On the contrary, it supports predictions that most stars begin their lives in a litter^{11,12}.

The projected separations between these bodies are of the order of 1,000–10,000 times the Earth–Sun distance, consistent with theoretical models¹¹. Significantly, these separations are smaller than the Jeans length in the cloud. This is the length scale on which the cloud would be expected to fragment on the basis of the competition between gravity and thermal pressure. The scale of fragmentation is a strong indication that turbulent motions in the cloud played a part, which is a significant boost to the turbulent-fragmentation model of star formation described above. Of course, the discovery of one system does not preclude other modes of binary formation.

In addition to its youth, this system is remarkable because it is convincingly gravitationally bound. In general, it is challenging to measure each of the quantities that go into the calculation for assessing gravitational

binding: the total mass of the system, the velocities of its members and their separations. The uncertainty in mass is introduced by unknown quantities such as the ratio of dust grains to hydrogen gas, or the ratio of the observed molecule, ammonia, to the total gas mass. Velocities and distances are both difficult to determine because we see only two-dimensional projections of three-dimensional quantities. Even folding in these uncertainties, Pineda and co-workers demonstrated that, as with Michell's double stars, these four are not just chance alignments.

Many questions remain about this system and others like it. We cannot predict whether the quadruplets will remain together or break up to form a triplet and a singlet. We do not know what the final orbital configuration will be, nor the ratio of the masses of the components. The ability to make these predictions, at least statistically, is crucial for connecting these infants with the descendants that conspire to explode as supernovae. And perhaps the biggest question for the field remains unanswered. What determines which star gets a sibling and which does not? The influx of data from the

current and next generation of telescopes such as the Atacama Large Millimeter/submillimeter Array and the James Webb Space Telescope will surely help to solve this puzzle. ■

Kaitlin M. Kratter is in the Department of Astronomy, University of Arizona, Tucson, Arizona 85721-0065, USA.

e-mail: kkratter@email.arizona.edu

1. Duquennoy, A. & Mayor, M. *Astron. Astrophys.* **248**, 485–524 (1991).
2. Raghavan, D. *et al.* *Astrophys. J. Suppl.* **190**, 1–42 (2010).
3. Pineda, J. E. *et al.* *Nature* **518**, 213–215 (2015).
4. Mathieu, R. D. *Annu. Rev. Astron. Astrophys.* **32**, 465–530 (1994).
5. Duchêne, G. & Kraus, A. *Annu. Rev. Astron. Astrophys.* **51**, 269 (2013).
6. Crutcher, R. M. *Annu. Rev. Astron. Astrophys.* **50**, 29–63 (2012).
7. Fisher, R. T. *Astrophys. J.* **600**, 769–780 (2004).
8. Offner, S. S. R. *et al.* *Astrophys. J.* **725**, 1485–1494 (2010).
9. Bate, M. R. *Mon. Not. R. Astron. Soc.* **419**, 3115–3146 (2012).
10. Bonnell, I. A., Bate, M. R., Clarke, C. J. & Pringle, J. E. *Mon. Not. R. Astron. Soc.* **323**, 785–594 (2001).
11. Reipurth, B. *et al.* in *Protostars and Planets VI* (eds Beuther, H., Klessen, R. S., Dullemond, C. P. & Henning, T.) 267–290 (Univ. Arizona Press, 2014).
12. Bodenheimer, P. *Annu. Rev. Astron. Astrophys.* **33**, 199–238 (1995).

question by exploring the nature of DNA sequences at the junction of two fused telomeres (repetitive DNA sequences that cap the ends of chromosomes). Such joining between the ends of different chromosomes is undesirable, and is normally suppressed. However, telomeres can fuse in the absence of protective mechanisms, including the absence of NHEJ⁹. When the authors sequenced mouse telomere fusion junctions in an NHEJ-deficient system, they observed seemingly random, permuted sequences attributable to MMEJ. They found that an enzyme called Polθ (encoded by the *Polq* gene in mice) promotes this process in mammalian cells.

Polθ is an atypical member of the DNA polymerase family of enzymes, which catalyse the synthesis of chains of nucleic acids. It contains both a helicase–ATPase domain, which separates the strands of a DNA helix using energy from the hydrolysis of ATP molecules, and an error-prone polymerase domain that can extend DNA strands from mismatched or unmatched termini. Previous work¹⁰ had demonstrated a role for Polθ in MMEJ in the fruit fly *Drosophila melanogaster*. Mateos-Gomez and colleagues' finding in mice therefore indicates that this role is evolutionarily conserved.

Intriguingly, the researchers show that MMEJ and HR are competing pathways, as evidenced by their observation of increased Rad51-dependent HR in cells lacking Polθ; Rad51 is a protein that performs HR-mediated repair. Furthermore, when the authors reduced Polθ levels in HR-deficient mouse cells harbouring mutations in either of the breast-cancer susceptibility genes

DNA REPAIR

Familiar ends with alternative endings

The faithful propagation of species requires a complex balance of DNA-repair pathways to maintain genome integrity. New work sheds light on one such poorly understood pathway and its role in certain cancers. SEE LETTERS P.254 & P.258

NAM WOO CHO & ROGER A. GREENBERG

The repair of double-strand breaks in DNA is crucial for life, and can proceed by several mechanisms. How are these repair pathways coordinated, and what happens when certain pathways are disabled? This question is especially pertinent because DNA-repair systems are deficient in many types of cancer; inhibition of complementary repair pathways^{1,2} might therefore be used to kill such cancer cells. Two papers^{3,4} in this issue, and an accompanying report⁵ published online in *Nature Structural and Molecular Biology*, reveal how a poorly understood alternative repair pathway compensates when the predominant mechanisms are impaired. The findings help to explain the causes of genomic rearrangements that are seen in cancer cells, and suggest therapeutic opportunities for cancers that harbour DNA-repair deficiencies.

Double-strand-break repair mostly occurs through mechanisms called homologous recombination (HR) and non-homologous

end-joining (NHEJ; Fig. 1). HR uses DNA ends that have been 'resected' to produce 3' single-stranded DNA overhangs, which probe for complementary strands — homologous DNA sequences — that act as templates for DNA repair. NHEJ does not require homologous DNA, and preferentially acts on 'blunt-ended' DNA breaks that do not have large overhangs.

But an alternative form of end-joining has been described^{6–8} that lies somewhere between HR and NHEJ. Most alternative end-joining seems to rely on microhomologies — short sequences of a few homologous base pairs — and thus is often referred to as microhomology-mediated end-joining (MMEJ). Unlike HR, MMEJ is inherently error-prone because the use of microhomology leads to repair events that produce deletions of sequences from the strand being repaired and rearrangement of sequences between pairs of chromosomes, known as translocations.

So how does MMEJ fit into the larger picture of double-strand-break repair? Mateos-Gomez *et al.*³ (page 254) have addressed this

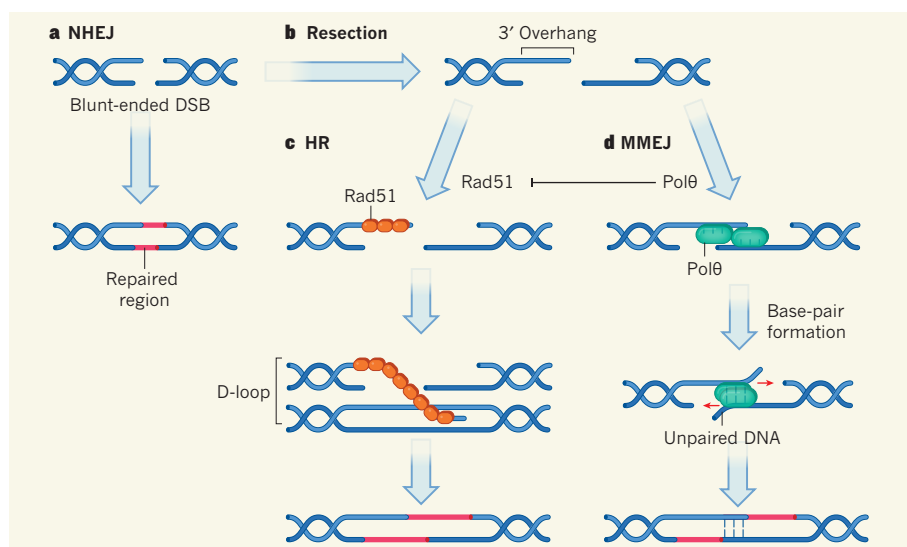


Figure 1 | Repair mechanisms for double-strand breaks in DNA. **a**, Blunt-ended double-strand breaks (DSBs) can be repaired by non-homologous end-joining (NHEJ). **b**, Resection of blunt ends generates 3' overhangs, allowing additional repair mechanisms. **c**, In homologous recombination (HR), further resection enables binding of Rad51 proteins, which help the overhang to bind to complementary (homologous) DNA in another DNA helix, forming a structure called a D-loop. The homologous DNA is used as a template for repair. **d**, A third mechanism, microhomology-mediated end-joining (MMEJ), is promoted by the enzyme Polθ. Kent *et al.*⁵ report that Polθ molecules bind to pairs of resected ends, enabling short homologous DNA sequences in overhangs to form base pairs, leaving adjacent regions unpaired. The enzyme then extends each strand (red arrows) from the base-paired region using the opposing overhang as a template, and the unpaired regions are removed. Mateos-Gomez *et al.*³ and Ceccaldi *et al.*⁴ find that Polθ also impedes HR by limiting Rad51 accumulation at resected ends. Both groups also show that HR-deficient cancers might rely on Polθ-mediated Rad51 inhibition and MMEJ for survival.

brca1 or *brca2* (or their human equivalents *BRCA1* and *BRCA2*), this greatly exacerbated chromosomal aberrancies and reduced the ability of the cells to survive, suggesting that Polθ-mediated MMEJ could compensate for the loss of HR. These findings provide enticing mechanistic insights into the observation¹¹ that cancer genomes involving *BRCA* mutants display characteristic signatures of MMEJ.

Ceccaldi and colleagues' study⁴ (page 258) began with the finding that Polθ is over-expressed in high-grade serous ovarian cancer, which is characterized by high rates of HR deficiency and frequent *BRCA* mutations. Like Mateos-Gomez *et al.*, they observed increased HR and RAD51 accumulation at DNA breaks in human cells depleted of Polθ, compared with normal cells. The authors also found that Polθ interacts directly with RAD51, and they identified several binding motifs for RAD51 on Polθ, one of which was necessary and sufficient for this interaction.

The researchers went on to show that both the ATPase- and RAD51-binding domains contribute to the inhibition of RAD51-dependent D-loop formation (a key step in the HR pathway), and to the inhibition of HR. Interestingly, they observed that Polθ-depleted cells have a mild hypersensitivity to DNA-damaging agents, and that this becomes much more striking when Polθ depletion is combined with HR deficiency caused by either depletion of FANCD2 (a protein involved in HR) or

BRCA1 mutation. In a clear example of genetic interaction, the investigators observed that mice died as embryos when both *Fancd2* and *Polq* genes were mutated, whereas mutation of either gene alone allowed the animals to survive to adulthood.

The two studies suggest that inhibition of Polθ might be therapeutically useful for HR-deficient cancers. But what aspect of Polθ should be inhibited? Ceccaldi *et al.* found that the RAD51-binding motifs are required to limit RAD51 accumulation at breaks and toxicity in HR-deficient cells, but that the polymerase domain is not essential. This suggests that the motifs are the appropriate target on Polθ. However, Mateos-Gomez *et al.* report high rates of HR in mouse embryonic stem cells expressing Polθ and in which the polymerase domain is inactivated by mutation, and find that these cells show reduced survival when *Brca1* expression is low. These observations suggest that the Rad51-binding motifs and the polymerase domain of Polθ may each contribute to survival of HR-deficient cells, although quantification of their relative contributions remains to be sorted out. Introducing defined mutations in each domain in an *in vivo* model could help to address this important issue.

Other fundamental questions are, how does Polθ perform MMEJ, and what are its substrates? Kent *et al.*⁵ show that human Polθ directly binds two ends of resected double-strand breaks in a process that depends on an evolutionarily conserved loop domain in the enzyme (Fig. 1d).



50 Years Ago

The Medical Research Council has set up a Brain Metabolism Research Unit in the Department of Pharmacology, University of Edinburgh Medical School. The Unit, which will be under the honorary direction of Prof. W. L. M. Perry, aims to undertake experimental and clinical studies of the metabolic pathways of certain amino-acids and other substances in the brain and tissue fluids. The action of psychotropic drugs on these pathways will be used as a means of trying to determine whether there are metabolic defects in the various psychoses and ultimately whether such defects can be corrected.

From *Nature* 13 February 1965

100 Years Ago

In answer to a question as to typhoid in the Army, asked in the House of Commons on February 8, Mr. Tennant, Under-Secretary of State for War, said:—"Of the 421 cases of typhoid in the present campaign among British troops 305 cases were in men who were not inoculated within two years. In the 421 cases there have been thirty-five deaths. Of these deaths thirty-four were men who had not been inoculated within two years. Only one death occurred among patients who were inoculated, and that man had only been inoculated once, instead of the proper number of times — namely, twice." This is a marvellous record; and no further answer than it provides is needed to the inhuman efforts made by anti-vaccinationists to induce men to object to inoculation ... we can only wonder at the patience of the British people in permitting a prejudiced faction to urge men not to subject themselves to a treatment by which they save others and themselves from suffering and death.

From *Nature* 11 February 1915

This allows microhomologous sequences in the overhangs to form base pairs. Pol θ then extends each strand from the base-paired region using the opposing overhang as a template. The researchers find that Pol θ requires partially resected DNA containing two to six base pairs of microhomology to perform MMEJ *in vitro*. This explains why MMEJ can act on substrates destined for HR, because both repair mechanisms require resection.

The three papers provide a beautiful illustration of how the multifunctional DNA-repair toolkit is coordinated to faithfully preserve genome integrity. They also reveal how cancer cells may thrive through increased usage of

error-prone repair pathways — a strategy that enables cellular survival by limiting the accumulation of persistent DNA damage, but at the expense of genome integrity. Understanding the mechanisms involved opens up targets for anticancer drug research. ■

Nam Woo Cho and Roger A. Greenberg are in the Departments of Cancer Biology and Pathology, Abramson Family Cancer Research Institute, Basser Research Center for BRCA, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104-6160, USA.
e-mail: rogergr@mail.med.upenn.edu

1. Bryant, H. E. *et al. Nature* **434**, 913–917 (2005).
2. Farmer, H. *et al. Nature* **434**, 917–921 (2005).
3. Mateos-Gomez, P. A. *et al. Nature* **518**, 254–257 (2015).
4. Ceccaldi, R. *et al. Nature* **518**, 258–262 (2015).
5. Kent, T., Chandramouly, G., McDevitt, S. M., Ozdemir, A. Y. & Pomerantz, R. T. *Nature Struct. Mol. Biol.* <http://dx.doi.org/10.1038/nsmb.2961> (2015).
6. Yan, C. T. *et al. Nature* **449**, 478–482 (2007).
7. Corneo, B. *et al. Nature* **449**, 483–486 (2007).
8. Boulton, S. J. & Jackson, S. P. *EMBO J.* **15**, 5093–5103 (1996).
9. Steir, A. & de Lange, T. *Science* **336**, 593–597 (2012).
10. Yu, A. M. & McVey, M. *Nucleic Acids Res.* **38**, 5706–5717 (2010).
11. Nik-Zainal, S. *et al. Cell* **149**, 979–993 (2012).

This article was published online on 2 February 2015.

GEOCHEMISTRY

When carbon escaped from the sea

A record of boron isotopes in fossils of microscopic plankton provides fresh evidence that some ocean regions were a source of carbon dioxide to the atmosphere when Earth warmed at the end of the last ice age. [SEE LETTER P.219](#)

KATHERINE A. ALLEN

During the last ice age, northern continents groaned under the weight of vast ice sheets and the concentration of carbon dioxide in the atmosphere was about 50% lower than it is today¹. Directly after the ice age, during a period known as the glacial termination, CO₂ probably escaped from the ocean into the atmosphere, but direct evidence for this carbon transfer has been hard to find. On page 219 of this issue, Martínez-Botí *et al.*² identify areas of CO₂ degassing from the ocean during the termination. Their records provide clues to the role of the ocean in the dramatic transition from an icy planet to the milder climate of the modern world.

Today, the exchange of CO₂ between ocean and atmosphere varies by season and location. In an average year, tropical oceans (14° S to 14° N) are a strong source of CO₂ to the atmosphere, whereas the mid-latitudes (14–50° S and N) and North Atlantic contain strong sinks, and the Southern Ocean (south of 50° S) swings seasonally between source and sink, resulting in a net neutral effect for that region³. These patterns of gas exchange depend primarily on the difference between the partial pressure of CO₂ (p_{CO_2}) in the ocean and that of the atmosphere (partial pressure is a measure of the pressure generated by a component of a mixture of gases). This difference (called Δp_{CO_2}) may have varied in the past. In general, if the partial pressure of CO₂ in seawater is greater than that in air

(that is, if Δp_{CO_2} is positive), then seawater will lose CO₂ to the air, and vice versa.

Ice-core records indicate that, during the Last Glacial Maximum (about 23,000 to 19,000 years ago), the atmospheric CO₂ concentration was roughly 30% lower than during the preceding warm period⁴. Somehow, during the

descent into the last ice age, the CO₂ balance between ocean and atmosphere was tipped in favour of the ocean. After the ice age, CO₂ levels rose again, implying that the balance must have tipped back. These shifts in carbon storage could have been caused by physical, chemical or biological mechanisms — such as changes in ocean circulation, sea-ice cover and phytoplankton productivity — potentially working in concert.

To gain insight into these possibilities, palaeoceanographers turn to sea-floor sediments. Marine sediments represent a rich archive of ocean history, in that the remains of biological matter and their chemical composition reflect past biological activity and ocean environmental conditions. Despite much progress, it has been difficult to tease out the mechanism underlying past changes in ocean CO₂ storage. A particular challenge has been the lack of direct evidence for exchange of CO₂

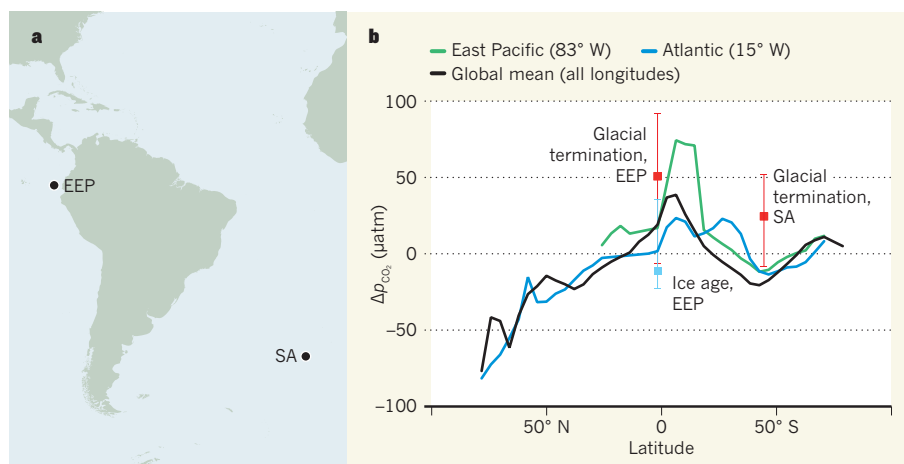


Figure 1 | Exchange of carbon dioxide between the ocean and atmosphere. **a**, Martínez-Botí *et al.*² have used boron isotopes from planktonic fossils to determine past differences between the partial pressures of CO₂ in the ocean and the atmosphere (Δp_{CO_2} , measured in microatmospheres) in the eastern equatorial Pacific Ocean (EEP) and the South Atlantic Ocean (SA). **b**, They find that Δp_{CO_2} in the EEP was negative, on average, during the latter part of the last ice age (25,000 to 19,000 years ago), implying that the ocean absorbed CO₂ from the atmosphere; ice-age data for the SA are not available. During the glacial termination (about 18,000 to 11,000 years ago), average Δp_{CO_2} at both sites was positive. This suggests that the ocean released CO₂ to the atmosphere at both locations at that time, which probably contributed to the overall rise in atmospheric CO₂ levels between the ice age and the following, milder Holocene epoch. Data for the modern ocean (solid lines) are from ref. 3. Squares represent mean values; vertical bars represent the full range of values for each period.

across the air–sea interface. Martínez-Botí *et al.* offer an insight into this exchange that is based on the boron isotopic composition ($\delta^{11}\text{B}$) of calcite — a form of calcium carbonate — produced by planktonic foraminifera.

Planktonic foraminifera are single-celled organisms that spend their lives floating in water. Many species inhabit the upper sunlit layers of the ocean, just below the air–sea interface, where CO_2 exchange occurs. Modern calibrations have shown that, as the pH of seawater increases, the $\delta^{11}\text{B}$ of calcite produced by planktonic foraminifera also increases^{5,6}.

Seawater pH and p_{CO_2} are also directly linked. This fact has enabled the atmospheric CO_2 record to be calculated⁷ from the $\delta^{11}\text{B}$ of calcite in sedimentary cores taken from beneath a mid-ocean gyre (a large rotating system of wind-driven ocean currents), where the ocean is roughly in equilibrium with the atmosphere (Δp_{CO_2} is approximately 0). The close agreement between this $\delta^{11}\text{B}$ -based CO_2 record and the record measured from bubbles of ancient air trapped in Antarctic ice demonstrated that $\delta^{11}\text{B}$ from the mid-ocean gyre is closely linked to atmospheric CO_2 across ice-age cycles.

By contrast with that study, Martínez-Botí *et al.* have targeted sedimentary cores from ocean upwelling regions to identify specific locations and periods of ocean CO_2 degassing (positive Δp_{CO_2}). This is not the first time that this has been done — the $\delta^{11}\text{B}$ record from the western equatorial Pacific Ocean⁸ suggested that the tropics may have been a significant source of CO_2 across the last glacial termination. The present study builds on that work by targeting regions in which this technique has not yet been applied: the eastern equatorial Pacific Ocean and the edge of the Southern Ocean in the South Atlantic Ocean.

The authors derived Δp_{CO_2} records by combining estimates of the partial pressure of CO_2 in the ocean with CO_2 data from ice cores. The records suggest that both the eastern equatorial Pacific and the South Atlantic were major sources of atmospheric CO_2 during the deglaciation (about 18,000 to 11,000 years ago; Fig. 1). Some of the highest Δp_{CO_2} values are accompanied by regional peaks in opal flux (the silica produced by marine microorganisms; see Fig. 3 of the paper). This is consistent with the idea that intensified ocean upwelling brought both nutrients and respired CO_2 to the sea surface during the deglaciation⁹.

Even more intriguingly, the record for the eastern equatorial Pacific does not follow the same pattern as that for the South Atlantic, implying that gas exchange in the two regions might be controlled by different processes. Together, these records imply that ventilation of CO_2 from the Southern Ocean increased the carbon content of the atmosphere during the last deglaciation, and that gas exchange in the tropics may also have had a pivotal role.

These exciting hints regarding ocean–atmosphere carbon exchange raise further

questions. Martínez-Botí and colleagues report that Δp_{CO_2} in the eastern equatorial Pacific fluctuated during the Last Glacial Maximum when atmospheric CO_2 levels were relatively stable, implying that other ocean regions may have acted as sinks. Records from other ocean regions are needed to confirm this. It should also be noted that, although Δp_{CO_2} records from certain key sites are invaluable for testing ice-age hypotheses, the data are not sufficient to quantify net global air–sea CO_2 exchange across ice-age cycles. For perspective, the uncertainty associated with estimates of modern CO_2 flux is $\pm 50\%$, even though there have been more than 3 million direct CO_2 measurements³. Still, the new results are promising, and their usefulness will increase as more records are established.

Additional records are now needed to cover the other half of the glacial cycle: the descent into the ice age. A consideration of Δp_{CO_2} obtained from the remains of planktonic foraminifera, along with proxies of deep-ocean carbonate chemistry (such as $\delta^{11}\text{B}$ or the ratio of boron to calcium in the fossils of

sediment-dwelling benthic foraminifera) may help palaeoceanographers to link sites of CO_2 storage in the interior ocean with exchange sites at the sea surface, thereby providing a deeper understanding of the link between carbon and climate. ■

Katherine A. Allen is in the Department of Marine and Coastal Sciences, Rutgers University, New Brunswick, New Jersey 08901, USA.
e-mail: kat.allen@rutgers.edu

1. Monnin, E. *et al.* *Science* **291**, 112–114 (2001).
2. Martínez-Botí, M. A. *et al.* *Nature* **518**, 219–222 (2015).
3. Takahashi, T. *et al.* *Deep Sea Res. II* **56**, 554–577 (2009).
4. Petit, J. R. *et al.* *Nature* **399**, 429–436 (1999).
5. Sanyal, A., Bijma, J., Spero, H. J. & Lea, D. W. *Paleoceanography* **16**, 515–519 (2001).
6. Henehan, M. J. *et al.* *Earth Planet. Sci. Lett.* **364**, 111–122 (2013).
7. Hönisch, B. & Hemming, N. G. *Earth Planet. Sci. Lett.* **236**, 305–314 (2005).
8. Palmer, M. R. & Pearson, P. N. *Science* **300**, 480–482 (2003).
9. Anderson, R. F. *et al.* *Science* **323**, 1443–1448 (2009).

NEURODEGENERATION

Cold shock protects the brain

A protein released during hypothermia has been found to affect the progression of neurodegenerative disease in mice by sparing neurons from death and preserving the connections between them. [SEE LETTER P.236](#)

GRAHAM KNOTT

History is littered with examples of the use of cold as a therapeutic agent — from the ancient Greek physician Claudius Galen, who used it to treat fevers, to its application for improving the outcomes of surgery after serious battlefield injuries. Mounting evidence of its effectiveness led surgeons in the 1950s to use hypothermia as a tool to ameliorate the side effects of brain surgery¹, and even today, cold is an effective treatment for birth asphyxia — a lack of oxygen during the perinatal period that can lead to brain damage. It was not until 1987 that an animal study² showed that cooling could reduce neuronal death after brain injury. In this issue, Peretti *et al.*³ (page 236) provide compelling evidence that a specific protein affects the brain's ability to reshape its connections in response to cooling, and that the protein's overexpression might provide therapeutic benefits in neurodegenerative disease.

The neuroprotection offered by cold has spurred efforts to understand the mechanisms

responsible, and findings so far offer insight into how brain metabolism and cell death are affected⁴. One aspect of neuroprotection research that is attracting interest involves a small group of proteins whose synthesis increases during hypothermia, even as production of other proteins decreases. These cold-shock proteins, which are found in many species, bind with RNA and mediate overall protein production. One such protein in particular, RNA-binding motif protein 3 (RBM3), is emerging as a central player in the protection of neurons after periods of hypothermia⁵. Investigations have focused mainly on how RBM3 and other cold-shock proteins affect the outcome of various types of stroke.

In their study, Peretti and colleagues subjected mice modelling Alzheimer's disease and mice infected with prions (proteins that cause neurodegenerative diseases such as Creutzfeldt–Jakob disease) to deep hypothermia, with the animals' body temperatures being reduced from 37 °C to 16–18 °C for 45 minutes. The authors identified and counted the synaptic connections between



ROGER TIDMAN/EPA

Figure 1 | A hibernating hazel dormouse. Peretti *et al.*³ show that a laboratory-based model of hibernation reveals extensive remodelling of synaptic connections, providing clues to what happens in the wild.

neurons using electron microscopy, and found that, during hypothermia, the number of contacts dropped significantly in a region of the brain implicated in memory formation — the hippocampus. After young animals had been warmed up, the number of synapses returned to normal.

In older animals, the same loss of synapses occurred, but the recovery did not. The effects on RBM3 also differed in old and young animals: whereas levels of RBM3 rose in response to the temperature drop in young mice, they failed to do so in older ones. Although the researchers found no symptoms of neurodegeneration in older animals on the basis of histological sections, biochemical measurements or behaviour, these mice were presumably further along the road to disease.

Neurodegenerative disease is known to involve the progressive loss of synapses and neurons, with a resulting reduction in cognitive ability. Peretti and co-workers' experiment is interesting because of the lack of response of RBM3 to cooling in pre-symptomatic older animals. Another aspect of note is that this is the first demonstration of synaptic loss followed by recovery using a laboratory-based model of hibernation. Previous work⁶ has shown some synaptic remodelling in hibernating animals that is suggestive of a weakening of connections between neurons; these may then strengthen on arousal. The current experiment, however, shows a more-extensive alteration in neuronal connectivity in laboratory mice. It would be useful to see how these synaptic changes alter the functioning of neuronal circuits in the hippocampus and other brain

regions during hibernation (Fig. 1).

The fact that plasticity and recovery of synaptic connections is missing when RBM3 levels do not rise led Peretti and colleagues to investigate whether artificially boosting levels of this molecule would protect against neurodegenerative disease. They therefore subjected prion-infected mice to two hypothermia treatments spaced one week apart. This caused RBM3 levels to remain raised for as long as eight weeks in young mice, and resulted in no synaptic or neuronal loss, or deficits in behaviour, well into what would normally be the terminal phase of the disease. Protection did not occur in older mice, nor, perhaps more importantly, when the researchers inhibited RBM3. The latter result directly implicates the RBM3 protein, rather than some other factor of the hypothermia treatment, in neuroprotection.

The authors then left the hypothermia approach behind, and instead merely raised the levels of RBM3 by injecting prion-infected mice with a viral construct containing this protein, causing cells to overproduce it. This again imparted neuroprotection and suppression of behavioural deficits.

Peretti and colleagues' work points to one way in which the progression of devastating degenerative diseases might be slowed or halted. But it also indicates that tight control of protein-synthesis pathways is key for maintaining neuronal circuitry in healthy brains. The authors found that when they decreased the levels of RBM3 in control animals, the density of synapses also fell. Although this result is given only cursory mention in the paper, it suggests that RBM3, which is present

at high levels when neuronal circuits are being laid down during brain development, is also involved in their maintenance later on. Because some synapses in the adult brain are constantly being lost and re-formed, this is perhaps not surprising. Synaptic turnover is crucial to the ability of neuronal circuits to modify their function in response to changes in activity⁷. This raises a question about the extent to which molecules such as RBM3 are involved in adult brain plasticity.

The authors' work also highlights the need for a more-complete picture of the complex mechanisms of neuronal protein production, particularly at the synapse. This could have far-reaching consequences, not only for how we tackle neurodegenerative diseases and brain lesions caused, for example, by stroke, but also for our understanding of, and approach to, disorders of brain plasticity such as ageing and mental retardation. ■

Graham Knott is at the Faculty of Life Sciences, EPFL, 1015 Lausanne, Switzerland.
e-mail: graham.knott@epfl.ch

1. Wang, H., Olivero, W., Wang, D. & Lanzino, G. *Acta Neurochir.* **148**, 565–570 (2006).
2. Busto, R. *et al.* *J. Cereb. Blood Flow Metab.* **7**, 729–738 (1987).
3. Peretti, D. *et al.* *Nature* **518**, 236–239 (2015).
4. Yenari, M. A. & Han, H. S. *Nature Rev. Neurosci.* **13**, 267–278 (2012).
5. Chip, S. *et al.* *Neurobiol. Dis.* **43**, 388–396 (2011).
6. Popov, V. I. & Bocharova, L. S. *Neuroscience* **48**, 53–62 (1992).
7. Holtmaat, A. & Svoboda, K. *Nature Rev. Neurosci.* **10**, 647–658 (2009).

This article was published online on 14 January 2015.

From quantum matter to high-temperature superconductivity in copper oxides

B. Keimer¹, S. A. Kivelson², M. R. Norman³, S. Uchida⁴ & J. Zaanen⁵

The discovery of high-temperature superconductivity in the copper oxides in 1986 triggered a huge amount of innovative scientific inquiry. In the almost three decades since, much has been learned about the novel forms of quantum matter that are exhibited in these strongly correlated electron systems. A qualitative understanding of the nature of the superconducting state itself has been achieved. However, unresolved issues include the astonishing complexity of the phase diagram, the unprecedented prominence of various forms of collective fluctuations, and the simplicity and insensitivity to material details of the ‘normal’ state at elevated temperatures.

The discovery of high-temperature superconductivity in the copper oxide perovskite $\text{La}_{2-x}\text{Ba}_x\text{CuO}_4$ (ref. 1) ranks among the major scientific events of the twentieth century. The superconducting transition temperatures in the copper oxides greatly exceed those of any previously known superconductor by almost an order of magnitude; in 1986 the highest possible temperature at which superconductivity could survive was widely believed to be 30 K (Fig. 1). Moreover, according to the theory of ‘conventional’ superconductors, the copper oxides would have seemed the least likely materials in which to look for superconductivity: at room temperature they are such poor conductors that they can hardly be classified as metals and, indeed, if their chemical composition is very slightly altered they become highly insulating antiferromagnets. Magnetism arises from strong repulsive interactions between electrons, whereas conventional superconductivity arises from induced attractive interactions, making magnetism and superconductivity seemingly antithetical forms of order.

The Bardeen–Cooper–Schrieffer (BCS) theory² of the late 1950s provided an extremely successful framework within which to understand conventional superconductors, and gave rise to conceptual breakthroughs. The basic insight is that the electrons collectively bind into ‘Cooper’ pairs and simultaneously condense in much the same way as bosons condense into a superfluid state. Fundamental to the BCS mechanism is the fact that, despite the strong direct Coulomb repulsions, the relatively weak attractions between electrons induced by the coupling to the vibrations of the lattice (phonons) can bind the electrons into pairs at energies smaller than the typical phonon energy. This was widely believed to imply that the superconducting transition temperature T_c of conventional superconductors could never exceed 30 K (ref. 3), although this limit has been revised upwards by the discovery in 2001 of superconductivity with $T_c = 39$ K in the simple metal MgB_2 (ref. 4), where circumstances conspire to optimize the electron–phonon mechanism. However, this is still far below the maximum T_c of the copper oxides.

As the properties of the copper oxides were studied with ever-increasing precision and sensitivity, it became clear that much of the well-understood quantum theory of the electronic properties of solids, which has been spectacularly successful in accounting for the properties of conventional metals and superconductors, fails entirely to address many features of the copper oxides and, more generally, of a broad array of ‘highly correlated electron systems’ of which the copper oxides are the most studied. (A schematic phase diagram of the copper oxides is shown in Fig. 2.)

Most prominently, at temperatures well above T_c the conductivity in the copper oxides is almost two orders of magnitude smaller than in simple metals and exhibits frequency and temperature dependences that are incompatible with the conventional theory of metals; this has led to materials in the regime above T_c being referred to as ‘strange metals’ or ‘bad metals’. The behaviour exhibited by these ‘strange metals’, much of which is simple to describe in terms of the so-called ‘marginal-Fermi-liquid phenomenology’⁵, has resisted any generally accepted understanding. On the other hand, similar behaviour has now been documented in a large number of electronically interesting materials⁶, indicating that this is a general property of strongly correlated electron systems, and is not directly linked to high-temperature superconductivity. We consider this to

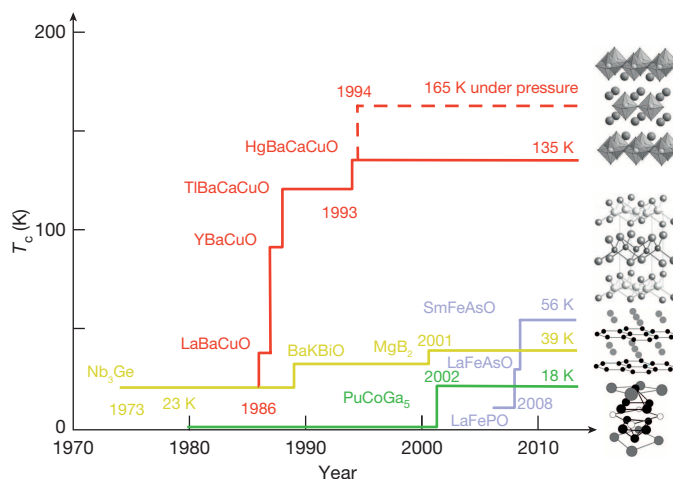


Figure 1 | T_c versus time. Superconducting transition temperatures versus year of discovery for various classes of superconductors. The images on the right are the crystal structures of representative materials. The established record for conventional electron–phonon superconductors (yellow) is 39 K in MgB_2 . Given the small Fermi energies, the T_c values found in the family of heavy fermion superconductors (green) are actually remarkably high. There has been much interest in recent years in the new family of ‘iron superconductors’ (purple) in which T_c values approach 60 K. The record holders are found in the copper oxide family (red), with a maximum T_c of 165 K found in a ‘mercury’ copper oxide under pressure (dashed red line).

¹Max Planck Institute for Solid State Research, Heisenbergstrasse 1, D-70569 Stuttgart, Germany. ²Department of Physics, Stanford University, Stanford, California 94305, USA. ³Materials Science Division, Argonne National Laboratory, Argonne, Illinois 60439, USA. ⁴Department of Physics, University of Tokyo, Bunkyo-ku, Tokyo 113-0033, Japan. ⁵Lorentz Institute for Theoretical Physics, Universiteit Leiden, PO Box 9506, 2300 RA Leiden, The Netherlands.

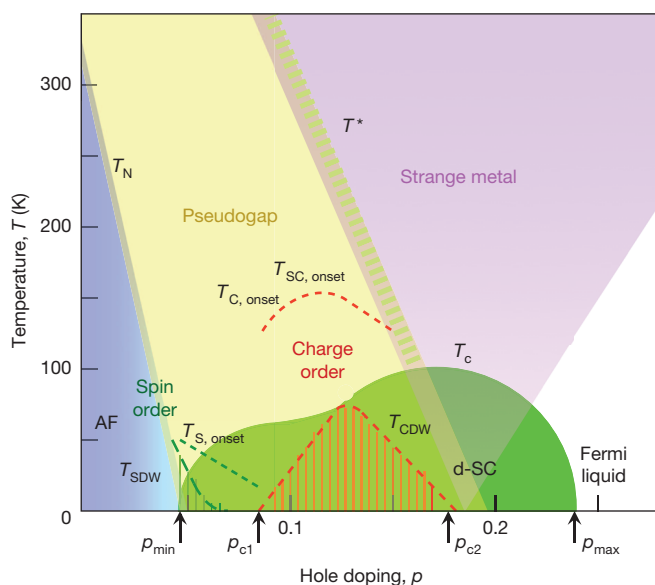


Figure 2 | Phase diagram. Temperature versus hole doping level for the copper oxides, indicating where various phases occur. The subscript ‘onset’ marks the temperature at which the precursor order or fluctuations become apparent. $T_{S, \text{onset}}$ (dotted green line), $T_{C, \text{onset}}$ and $T_{SC, \text{onset}}$ (dotted red line for both) refer to the onset temperatures of spin-, charge and superconducting fluctuations, while T^* indicates the temperature where the crossover to the pseudogap regime occurs. The blue and green regions indicate fully developed antiferromagnetic order (AF) and d -wave superconducting order (d-SC) setting in at the Néel and superconducting transition temperatures T_N and T_c , respectively. The red striped area indicates the presence of fully developed charge order setting in at T_{CDW} . T_{SDW} represents the same for incommensurate spin density wave order. Quantum critical points for superconductivity and charge order are indicated by the arrows.

be the most important open problem in the understanding of quantum materials, and it is here that radically new ideas, including those derived from recently developed non-perturbative studies in string theory, may be useful.

More unique to the copper oxides is the behaviour observed in a range of temperatures immediately above T_c in what is referred to as the ‘pseudogap’ regime. It is characterized by a substantial suppression of the electronic density of states at low energies that cannot be simply related to the occurrence of any form of broken symmetry. Although much about this regime is still unclear, convincing experimental evidence has recently emerged that there are strong and ubiquitous tendencies towards several sorts of order or incipient order, including various forms of charge-density-wave, spin-density-wave, and electron-nematic order. There is also suggestive, but far from definitive, evidence of several sorts of novel order—that is, never before documented patterns of broken symmetry—including orbital loop current order and a spatially modulated superconducting phase referred to as a ‘pair-density wave’. There are many fascinating aspects of these ‘intertwined orders’ that remain to be understood, but their existence and many aspects of their general structure were anticipated by theory⁷. Superconducting fluctuations also have an important role in part of this regime, although to an extent that is still much debated.

The high-temperature superconducting phase itself has a pattern of broken symmetry that is distinct from that of conventional superconductors. Unlike in conventional s -wave superconductors, the superconducting wavefunction in the copper oxides has d -wave symmetry^{8,9}, that is, it changes sign upon rotation by 90° . Associated with this ‘unconventional pairing’ is the existence of zero energy (gapless) quasiparticle excitations at the lowest temperatures, which make even the thermodynamic properties entirely distinct from those of conventional superconductors (which are fully gapped). The reasons for this, and its relation to a proximate antiferromagnetic phase, are now well understood, and indeed were also anticipated early on by some theories^{10–12}. However, while various attempts

to obtain a semiquantitative estimate of T_c have had some success¹³, there are important reasons to consider this problem still substantially unsolved.

Highly correlated electrons in the copper oxides

The chemistry of the copper oxides amplifies the Coulomb repulsions between electrons. The two-dimensional copper oxide layers (Fig. 3) are separated by ionic, electronically inert, buffer layers. The stoichiometric ‘parent’ compound (Fig. 2, zero doping) has an odd-integer number of electrons per CuO_2 unit cell (Fig. 3). The states formed in the CuO_2 unit cells are sufficiently well localized that, as would be the case in a collection of well-separated atoms, it takes a large energy (the Hubbard U) to remove an electron from one site and add it to another. This effect produces a ‘traffic jam’ of electrons¹⁴. An insulator produced by this classical jamming effect is referred to as a ‘Mott insulator’¹⁵. However, even a localized electron has a spin whose orientation remains a dynamical degree of freedom. Virtual hopping of these electrons produces, via the Pauli exclusion principle, an antiferromagnetic interaction between neighbouring spins. This, in turn, leads to a simple (Néel) ordered phase below room temperature, in which there are static magnetic moments on the Cu sites with a direction that reverses from one Cu to the next^{16,17}.

The Cu-O planes are ‘doped’ by changing the chemical makeup of interleaved ‘charge-reservoir’ layers so that electrons are removed (hole-doped) or added (electron-doped) to the copper oxide planes (see the horizontal axis of Fig. 2). In the interest of brevity, we will confine our discussion to hole-doped systems. Hole doping rapidly suppresses the antiferromagnetic order. At a critical doping of p_{min} , superconductivity sets in, with a transition temperature that grows to a maximum at p_{opt} , then declines for higher dopings and vanishes for p_{max} (Fig. 2). Materials with $p < p_{\text{opt}}$ are referred to as underdoped and those with $p_{\text{opt}} < p$ are referred to as overdoped.

It is important to recognize that the strong electron repulsions that cause the undoped system to be an insulator (with an energy gap of 2 eV) are still the dominant microscopic interactions, even in optimally doped copper oxide superconductors. This has several general consequences. The resulting electron fluid is ‘highly correlated’, in the sense that for an electron to move through the crystal, other electrons must shift to get out of its way. In contrast, in the Fermi liquid description of simple metals, the quasiparticles (which can be thought of as ‘dressed’ electrons) propagate freely through an effective medium defined by the rest of the electrons. The failure of the quasiparticle paradigm is most acute in the ‘strange metal’ regime, that is, the ‘normal’ state out of which the pseudogap and the superconducting phases emerge when the temperature is lowered. Nonetheless, in some cases, despite the strong correlations, an emergent Fermi liquid arises at low temperatures. This is especially clear in the overdoped regime (Fig. 2). But recently it has been shown that even in underdoped materials, at temperatures low enough to quench superconductivity by the application of a high magnetic field, emergent Fermi liquid behaviour

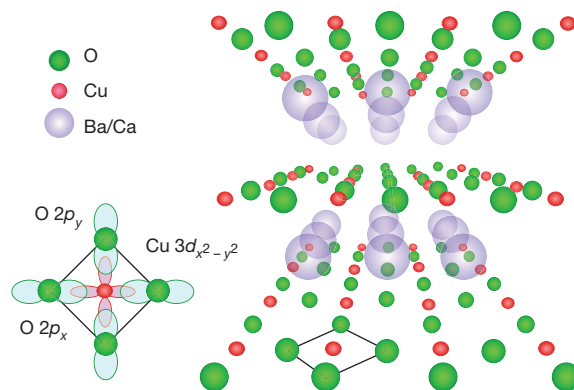


Figure 3 | Crystal structure. Layered copper oxides are composed of CuO_2 planes, typically separated by insulating spacer layers. The electronic structure of these planes primarily involves hybridization of a $3d_{x^2-y^2}$ hole on the copper sites with planar-coordinated $2p_x$ and $2p_y$ oxygen orbitals.

arises, albeit with characteristics (for example, a reconstructed Fermi surface) that are quite different from those predicted by band theory^{18,19}. Nevertheless, over most of the phase diagram, the frustration of the coherent electron motion produces physics that is qualitatively distinct from that of simple metals.

Although the large zero-point energy of electrons in a usual metal results in a quantum ‘rigidity’ that greatly suppresses all forms of inhomogeneous states, the Mott physics and the short-range antiferromagnetic correlations inherited from the undoped ‘parent’ compound combine to produce a local tendency to phase separation and various forms of order, which spontaneously break the translational symmetry of the underlying crystal^{20–22}. Thus, especially in the pseudogap regime of the phase diagram, it is unsurprising that various forms of order occur on intermediate length scales.

Pairing in an unconventional superconductor

It is now well established that electrons can form pairs, even when they repel each other at a microscopic scale. However, this involves non-trivial physics. A model that is often used as a point of departure for theoretical discussions is the famous Hubbard model, describing electrons hopping on a lattice parametrized in terms of the bandwidth $W = 8t$ (where t is a measure of the ‘hopping’ energy gain due to delocalization of the electrons) and an on-site electron–electron repulsion U . In the copper oxides, U and W are comparable. Even for this simplified model, analytic solutions are not available. However, approximate solutions of the doped Hubbard model can be obtained in several ways, and these invariably point to a d -wave superconducting ground state.

An intuitive understanding of the mechanism of pairing is best obtained by approaching the problem from an unrealistic weak-coupling perspective, that is, assuming $U \ll W$ (ref. 23). Here, the gap structure is determined by the solution of a variant of the original BCS equations, in which an appropriately renormalized two-particle vertex function, $\Gamma(\mathbf{k})$, plays the part of an effective interaction. For the case of purely repulsive interactions, if Γ is sufficiently \mathbf{k} -dependent, a sign-changing superconducting order parameter (where $\Delta(\mathbf{k})$ and $\Delta(\mathbf{k} + \mathbf{Q})$ have opposite sign) results for which interactions involving small momentum transfer are pair-breaking, and those with large momentum transfer near \mathbf{Q} promote pairing. In particular, if there are antiferromagnetic correlations, this typically implies a peak in Γ at the antiferromagnetic ordering vector, $\mathbf{Q} = \mathbf{Q}_{AF}$ (ref. 24), which is also an ideal vector for scattering between ‘antinodal’ regions of the Fermi surface of the copper oxides shown in Fig. 4; that is, precisely those regions where the d -wave gap is largest and of opposite sign. The gap ‘nodes’ along the diagonals of the Brillouin zone are then, in turn, where the d -wave gap vanishes.

Superconductivity in the Hubbard model cannot truly be approached from the strong-coupling limit, since there is now strong numerical evidence that for a broad range of doping, the ground state of the Hubbard model is ferromagnetic rather than superconducting for large enough U/t (ref. 25). However, the closely related t - J model (with $J = 4t^2/U$ being the superexchange interaction between copper spins mediated by the intervening oxygen ions) incorporates the essence of the strong-coupling physics through the constraint that no more than one electron at a time can occupy a given site. The t - J model can then be addressed, with values of $J/t \approx 0.5$, as a reasonable model in its own right. Although no controlled solution is known, the superconducting tendencies of this model have been investigated numerically since the early days of high-temperature superconductivity research^{26,27}. It is striking that the character and symmetry of the superconducting state itself and its association with short-range antiferromagnetic correlations look grossly similar, regardless of perspective²⁴.

Although intermediate coupling problems have thus far not been successfully solved by controlled analytic approaches, the lack of any small dimensionless parameters probably implies the lack of any long emergent length scales in the problem, except near a quantum critical point (QCP). With this in mind, a variety of numerical techniques have been developed to study this regime, including exact diagonalization (limited to small clusters), quantum Monte Carlo and its derivatives (variational

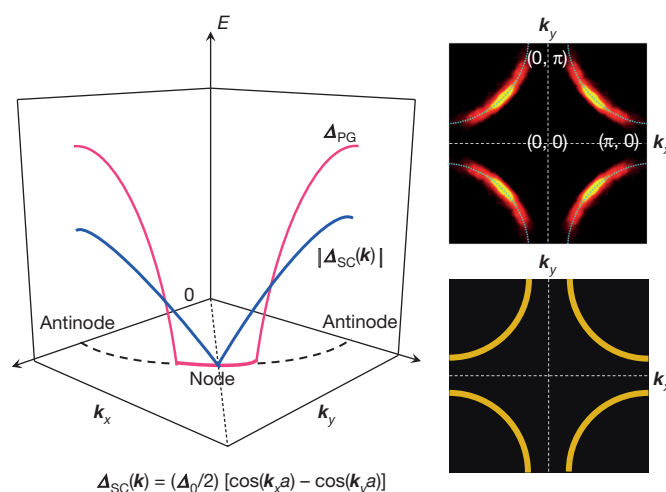


Figure 4 | Fermi surface, Fermi arcs and gap functions. The large Fermi surface predicted by band theory is observed by ARPES and STS for overdoped compounds (bottom right). But once the pseudogap sets in, the antinodal regions of the Fermi surface near the Brillouin zone edge are gapped out, giving rise to Fermi arcs (top right). This is reflected (left) in the angle dependence of the energy E of the superconducting gap Δ_{SC} (blue line) and pseudogap Δ_{PG} (red line) as functions of the momenta k_x and k_y in one quadrant of the Brillouin zone around the underlying large Fermi surface (dashed curve), as revealed by ARPES and STS. Note the gapless region around the d -wave superconducting node for the pseudogap case that defines the Fermi arcs. These arcs appear to be reconstructed into electron pockets centred at $(Q/2, Q/2)$ once charge order sets in, as revealed by quantum oscillation studies, where $(Q, 0)$ is the charge order wavevector¹⁹.

and fixed node approximations to get around the issue of negative probabilities in fermion simulations), dynamical mean field theory²⁸ and its cluster generalizations (either in momentum space or real space), density matrix renormalization group (designed for one-dimensional problems but can simulate strips), and its two-dimensional generalizations²⁹. These methods all have their pros and cons. They have, however, taught us that if superconductivity occurs, it is invariably of d -wave symmetry, but also that many competing states are close in energy, especially unidirectional charge order^{30,31}.

High- T_c superconductivity

Of course, there is additional complexity in going from theoretical results for simple model problems to the real experimental systems. Static antiferromagnetism disappears quickly as a function of doping (Fig. 2), but both inelastic neutron scattering and resonant inelastic X-ray scattering reveal that the antiferromagnetism of the insulator survives in the superconductor to a degree in the form of dynamical magnetic fluctuations which are much stronger than in conventional metals (and are strongly renormalized when cooling below T_D)^{32–35}. It is physically appealing to use the measured spin fluctuation spectrum as an approximation of the vertex Γ mentioned above. This yields reasonable values for T_c , and also appears to be consistent with some of the single electron self-energy effects detected by various electron spectroscopies such as ARPES (angle resolved photoemission spectroscopy) and STS (scanning tunnelling spectroscopy)^{13,36}.

However, this approach, departing from the idea of the attractive force (‘glue’) induced by spin fluctuations, has shortcomings³⁷. Despite its intuitive appeal, it is not based on controlled mathematics, since the same electrons that are pairing also form the ‘glue’. Another difficulty is that these simplified models leave out other effects that can influence the magnitude of T_c . A case in point is the electron–phonon interaction. There is good evidence that phonons affect both ARPES and STS line shapes³⁸, while strong anomalies are seen in the phonon spectra³⁹. There are a number of other neglected effects that are worrisome, in particular the non-local Coulomb interaction which is an especially relevant concern given the poor screening in the direction perpendicular to the planes⁴⁰.

More seriously, some fundamental aspects of high- T_c superconductivity are qualitatively different from the BCS variety. An example is the influence of quenched disorder. In absolute terms, most copper oxides can be regarded as chemically ‘dirty’, owing to their doped (non-stoichiometric) nature. In BCS theory, an important difference between s -wave and unconventional superconductivity is that the former is relatively impervious to structural disorder, while the latter is readily degraded⁴¹. The strong inhomogeneity seen by STS leaves no doubt that many copper oxide superconductors reside in a disordered lattice potential, but the d -wave superconductivity appears to be fairly insensitive to this adverse condition. It has been suggested that this is a consequence of strong local correlations⁴², but it still remains a puzzle.

A very basic quantity for the superconducting order is the superfluid density ρ_s , the quantity that parametrizes the rigidity of the phase of the superconducting order parameter, which also determines the capacity of the superconductor to expel electromagnetic fields. One can identify a temperature associated with the fluctuations of the phase $T_\theta \approx \rho_s/m^*$ where m^* is the effective mass. In a BCS superconductor, ρ_s is equal to the total density of electrons at zero temperature and accordingly $T_\theta \gg T_c \approx 0.5\Delta_0/k_B$, the temperature associated with the formation of the pairs where Δ_0 is the average superconducting gap (k_B is the Boltzmann constant). The fluctuations of the phase of the superconducting condensate are largely irrelevant; once Cooper pairs form, they automatically condense. Turning to the copper oxides, it was established early on that the superfluid density is anomalously small, scaling in the underdoped regime with T_c (the ‘Uemura law’⁴³). The conclusion is that in the underdoped copper oxides, T_θ and the pair-binding energy are of the same order and the thermal fluctuations of the phase should be crucial for the thermodynamics⁴⁴. A long-standing question is whether, perhaps, pairs already form at the (very high) pseudogap temperature T^* (Fig. 2), while at a much lower temperature, the actual T_c , the phase locks to form the long-range ordered superconducting state. As we will see, the physics of the phase fluctuations is intertwined with that of competing order.

In many cases, in copper oxides with the highest T_c (that is, at optimal doping), the superconductivity emerges directly as an instability of the strange metal phase. The strange metal is the least understood part of the phase diagram, because it does not appear to be describable in terms of weakly interacting Landau quasiparticles. The very non-BCS transition from the strange metal to the more conventional physics of the superconducting state is vividly apparent in the temperature evolution of the ARPES spectra at momenta near the ‘antinodes’ (Fig. 4) where the pairing forces of the d -wave superconductor are supposedly the strongest. For these momenta, the electron spectral function is strongly broadened as a function of energy⁴⁵. Upon entering the superconducting phase, a quasiparticle peak starts to develop that has the classic backbending (Bogoliubov) dispersion of a BCS superconductor⁴⁶. This is, in turn, consistent with the onset of coherence observed in microwave, infrared and thermal conductivities. But, unlike in BCS theory, it appears that the spectral weight of this antinodal ‘Bogoliubov’ is linearly proportional to the superfluid density, both as a function of doping and temperature^{47,48}. This is not understood: it is as if the phase coherence of the superconductor ‘freezes out’ single-particle coherence from the highly collective non-Fermi-liquid strange metal state.

Pseudogap regime

A prominent feature of this regime of the phase diagram is the line T^* , which denotes the onset of a partial gap observed in spectroscopic data. First inferred from nuclear magnetic resonance measurements that showed a reduction in the low-frequency spin excitations^{49,50}, this pseudogap was subsequently seen in c -axis polarized infrared conductivity measurements and is associated with a pronounced upturn in the c -axis resistivity⁵¹. In contrast, the in-plane polarized infrared conductivity indicates a drop in the scattering rate⁵², which is reflected in a reduction of the planar resistivity^{53,54}.

A much debated question is whether conventional (Hartree–Fock) mean-field treatments are able to provide even a qualitatively correct account of

the pseudogap phenomenology. There are surely issues related to the ‘plethora of orders’ discussed immediately below, something which is not easy to understand in the conventional way. However, one obtains a much sharper view using electron spectroscopies. The striking difference in the nature of the electronic excitations measured in ARPES when crossing from the ‘coherent’ nodal region to the ‘incoherent’ antinodal region in momentum space is called the nodal–antinodal dichotomy⁵⁵. The nodal region involves a narrow region around the zone diagonals, which gradually grows with increasing doping until it encompasses the entire Fermi surface in sufficiently overdoped materials. While the antinodal region lacks any quasiparticle-like spectral peaks, throughout the pseudogap regime it exhibits a suppression of low-energy electronic spectral weight on an energy scale that corresponds to the pseudogap⁴⁵.

The astonishing character of these observations is best illustrated by showing a map of the spectral weight at low energy as a function of \mathbf{k} in the first Brillouin zone (Fig. 4). In a Fermi liquid, the Fermi surface delineates the boundary between occupied and unoccupied quasiparticle states, so no matter how complicated it may be, the one thing it cannot do is abruptly end. However, in the pseudogap regime, there appear to be ‘Fermi arcs’ in the nodal regime⁵⁶. In a mean-field theory, the effective potential associated with a (density-wave) state that breaks translational symmetry can reconstruct a large Fermi surface, producing small Fermi surface pockets, but these must still form connected manifolds. It is plausible that the Fermi arcs are actually the front half of such a pocket^{57,58}, and hence there has been an intense search to find the ‘backside of the pocket’, but at present there is no definitive sign of it.

STS has proved particularly revealing in this context. Such data (mostly below T_c) exhibit electronic waves in real space that upon Fourier transformation show peaks that disperse with bias and have been mapped to scattering across the Fermi surface⁵⁹. One finds that in the superconducting state, the low-energy excitations near the nodes behave just as one would expect for a BCS d -wave state⁶⁰, but at higher energies, cross over to a dispersionless pattern characteristic of short-range stripe order. Interestingly, this low-energy dispersing pattern maps out the same Fermi arc⁶¹ in momentum space (Fig. 4) as observed directly by ARPES⁵⁶ in the pseudogap state, with the arc recovering the full Fermi surface once the doping exceeds a critical value⁶². However, the STS result, indicating a complete loss of coherence in the antinodal regime, appears to be inconsistent with the ARPES, which sees antinodal quasiparticles below T_c , even for underdoped materials⁶³. The higher-energy dispersionless pattern is seen at all energies when moving above T_c into the pseudogap state^{64,65}, consistent with local charge order, and has been identified as coexisting with the low-energy dispersing signal below T_c as well^{66,67}. But the consistency of the ARPES data with charge order is still an active area of debate⁶⁸—to date, no unambiguous signatures associated with the stripe wavevectors have been found.

Precursor pairing

The structure of the pseudogap in momentum space was directly mapped by ARPES experiments at temperatures between T^* and T_c , and found to crudely mimic the d -wave superconducting gap: the pseudogap is apparent only in the ‘antinodal’ regions of the Brillouin zone (Fig. 4)^{57,69,70} where the d -wave gap is largest. This immediately suggests that at the very high pseudogap temperature T^* , pairs already start to form, while phase fluctuations prohibit superconducting order until much lower temperatures are reached. So long as there is substantial short-range phase coherence, superconducting fluctuations should have large and identifiable signatures. For instance, for a range of temperatures that extends up to about $1.5T_c$ (but not to temperatures comparable to T^*), large fluctuation conductivity (at both direct and alternating current) is observed⁷¹, but there is some debate about how much such signatures differ from those observed in classic superconductors.

The more far-reaching notion of pairing correlations without substantial phase coherence persisting to temperatures of the order of T^* is difficult to define precisely, even in principle. The best circumstantial evidence comes from diamagnetism, which is observed up to about 150 K (ref. 72).

Though weak compared to full Meissner screening, it is still large compared to that of simple metals. Moreover, in underdoped $\text{YBa}_2\text{Cu}_3\text{O}_{6+x}$ (YBCO), a moderately well-defined interlayer Josephson plasma resonance seems to persist up to similar temperatures⁷³, and recent pump–probe experiments are consistent with transient superconducting order existing all the way to T^* (ref. 74). Perhaps the most compelling evidence is in the temperature evolution of the gap itself; the nodal–antinodal dichotomy notwithstanding, the pseudogap above T_c evolves remarkably smoothly into the gap of the superconducting state well below T_c . In that context, it has been suggested by ARPES that the Fermi arc is simply due to the lifetime broadening of a d -wave node^{75,76}, and this has been inferred as well by STS below T_c (ref. 77). The reconciliation of this superconducting-like signature in the fermion response in the pseudogap phase and an energy gap due to competing order⁷⁸ has been a major challenge, even more relevant now, given the new findings of such crystalline order (see below).

Competing orders

Another increasingly well-documented feature of the pseudogap regime is a tendency towards a variety of orders (that is, patterns of broken symmetry) in addition to superconductivity. Some involve ‘crystallization’ of the electrons in the form of stripes and other forms of charge order, but others appear to be more novel quantum liquids. There is also some evidence for new types of order involving various patterns of equilibrium orbital currents, and possibly a new sort of spatially modulated superconducting state.

Neutron scattering studies in the mid-1990s led to the experimental discovery of electronic ‘stripes’ in the $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$ (LSCO) family²². These studies were inspired, in part, by earlier mean-field theories of density-wave formation in lightly doped Hubbard-like models²⁰. An alternative and complementary theoretical perspective was based on the observation that doping an insulating antiferromagnet produces a tendency to phase separation that is frustrated by the long-range Coulomb interaction; the compromise is to form conducting stripe-like textures²¹. The stripe order, characterized by incommensurate antiferromagnetic order and charge segregation, was initially found in underdoped versions of LSCO, where a low-temperature tetragonal lattice deformation apparently acts as a pinning potential for the stripes²². However, it became clear that these stripes were different from the ‘classical’ stripes in the other doped Mott insulators: the copper oxide stripes stay metallic and even superconduct at low temperatures. Although the spatial organization looks similar to the mean field stripes, a crucial difference is that these can now be viewed as a partially crystallized superconductor, formed from electron pairs^{31,79}.

In general terms, a competition between superconductivity and crystallization is a very natural way to account for the diminishing superfluid density in the pseudogap regime. Indeed, quite recently, evidence has emerged that materials with static stripes form a “pair density wave”: the charge stripes are internally superconducting, but the phase reverses from stripe to stripe⁷. Given that the stripe orientation changes as one moves from one layer to the next, this frustrates the Josephson coupling between layers, entirely extinguishing the superfluid stiffness perpendicular to the planes and thus giving rise to a two-dimensional superconducting state consistent with transport measurements^{80,81}. If confirmed, this would constitute the discovery of a new phase of matter.

It was subsequently found, by inelastic neutron scattering, that the magnetic spectrum of the ordered stripes has a unique ‘hourglass’ pattern⁸², with the neck of the hourglass located at the commensurate antiferromagnetic wavevector, and this pattern has subsequently been observed in the insulating charge ordered states of manganites and cobaltates⁸³. Although this makes it natural to associate the inelastic neutron scattering spectrum with the spin waves associated with incommensurate magnetic order, in the copper oxides, this pattern persists for larger dopings, even where the stripe order is no longer condensed and where the character of the spin-wave spectrum changes dramatically when the temperature drops below T_c (ref. 84). In this regime, many salient features of the magnetic spectrum are similar to what is expected for weakly interacting quasiparticles

in a homogeneous d -wave superconductor⁸⁵. A reconciliation of these two very different pictures remains a challenge for the field.

For many years, static stripe order had seemed to be confined to the LSCO family. However, recently charge ordering was discovered in underdoped YBCO⁸⁶ and Bi- and Hg-based copper oxides^{67,68,87}. X-ray experiments find short-range incommensurate charge order that gradually sets in between 100 K and 200 K (refs 88 and 89). Moreover, high-resolution X-ray scattering⁹⁰ and nuclear magnetic resonance experiments have confirmed that the short-range charge order is truly static, and thus presumably arises from pinning of correlated charge fluctuations by defects⁹¹. Unlike in the stripes of the LSCO family, there is no evidence of coincident static (or nearly static) magnetic order. Moreover, the variation of the stripe wavevector with doping in YBCO⁹² and Bi-based copper oxides^{67,68} is much weaker and has the opposite sign of that in LSCO. Whereas in LSCO, this wavevector increases with doping as would be expected in a real space picture, in YBCO and the Bi-based copper oxides, the wavevector decreases, as would be expected from a momentum space picture involving vectors spanning the Fermi surface. This difference may be connected with differences in the spin behaviour: in YBCO a large spin gap is present that acts to suppress the incommensurate spin order that is more prevalent in the LSCO family^{34,35}.

Yet another interesting hint regarding the unusual relationship between the charge order and superconductivity follows from the temperature evolution of the charge order. The X-ray signal begins to build up smoothly upon cooling below a characteristic charge ordering temperature (T_{CDW}) typically less than T^* , to attain a maximum at the superconducting T_c , then drops noticeably below T_c , indicating competition between the charge order and superconductivity^{88,89,92}.

There is also evidence for ‘quantum nematic liquid crystal’ order occurring in the pseudogap phase. Such phases are translationally invariant but break point group (for example, rotational) symmetries. First suggested in the context of the quantum melting of stripe crystals⁹³, evidence appeared for a phase breaking the fourfold symmetry of the square lattice in underdoped YBCO, from transport^{94,95} and inelastic neutron scattering measurements⁹⁶. This ‘nematic’ signal was also found in an analysis of the STS data, showing that besides the ‘stripy’ texture breaking translations, there is also an overall (zero wavevector) breaking of rotations present, consistent with the two oxygen ions in the CuO_2 unit becoming inequivalent^{97,98}, though this is controversial⁹⁹.

These orders are all close siblings of the electron crystal. However, there is also evidence for a completely different kind of order below T^* as well. This order is symmetry-wise equivalent to having magnetic moments on the oxygen sites, and thus would be a magnetic analogue of the charge nematic mentioned above¹⁰⁰. But the original proposal that motivated the experiments involved spontaneous electron currents flowing inside the CuO_2 units in such a way that although rotational symmetry is broken, translational symmetry is not¹⁰¹. It has a quite distinct magnetic diffraction pattern, which was subsequently seen by spin-flip neutron scattering¹⁰⁰. If confirmed, this would again amount to the discovery of a new phase of matter, though it does not yield a natural explanation for the pseudogap, just by the very fact that it does not break translational symmetry. The real difficulty with this proposal is that current order should also be seen by local magnetic probes such as muon spin resonance and nuclear magnetic resonance, but this has not been observed. Potentially related to this is the onset of a small Kerr rotation at T^* , which also indicates some type of symmetry breaking¹⁰². This Kerr signal defines a phase line that cuts through the superconducting dome, vanishing near 18% doping.

The strange metal

The strange metal regime was recognized early on as perhaps the most mysterious aspect of the copper oxide phase diagram. The most basic difference between the strange metal and a conventional metal is the absence of quasiparticles. This has consequences for physical properties like the electrical resistivity. In a normal metal, unless the metal melts first, the resistivity saturates at high temperatures when the mean free path, l , becomes

of the order of the electron de Broglie wavelength, λ . The resistivity of the copper oxide strange metal can be linear in T from near T_c up to as high a temperature as measured¹⁰³, even when the inferred l would be smaller than λ , which would violate the Heisenberg uncertainty principle for the quasiparticles. Moreover, the Hall resistivity has a temperature dependence different from what would be expected in a quasiparticle picture¹⁰⁴.

In the late 1980s, some of these and various other experimental anomalies were encapsulated in the phenomenological ‘marginal Fermi liquid’ theory⁵. This asserts that the Fermi gas is coupled to a continuum of excitations that is spatially featureless, with a spectral density which is constant for $\omega > T$, but proportional to T for $\omega < T$. This leads to a damping rate that scales as $\max(\omega, T)$. This was confirmed later by high-resolution ARPES measurements, with the caveat that this is only seen in the nodal region, with the antinodal region behaving in a more incoherent fashion¹⁰⁵.

In the 1990s it was proposed that quantum criticality could explain the low-energy excitations of the strange metal. A quantum phase transition occurs when a continuous phase transition occurs at zero temperature as a function of a tuning parameter (like pressure or doping), where the corresponding QCP defines the boundary between the ordered (broken symmetry) and disordered quantum phases¹⁰⁶. The correlations at a QCP are characterized by a spatio-temporal scale invariance, which in turn has the effect that there are no longer quasiparticle poles (δ -functions) in spectral functions. Instead one finds power-law behaviour (‘branch cuts’) and spectral functions at finite temperature that are scaling functions of ω/T . This can be interpreted in terms of a dissipative energy relaxation time $\hbar/k_B T$, which is sometimes referred to as ‘Planckian dissipation’ because it is a quantum effect independent of material parameters¹⁰⁷. Moving away from the critical point, the energy scale above which scale invariance remains gradually increases. Accordingly, in the ‘tuning parameter’–temperature plane, there is a quantum critical wedge opening up from the QCP. This suggests an interpretation of the phase diagram in Fig. 2, where the strange metal is identified with the quantum critical wedge associated with a QCP under the superconducting dome near optimal doping.

The theory of quantum criticality in metallic systems is still a work in progress. One issue is that there may be reasons to believe that the QCP is intrinsically unstable, since the order parameter fluctuations mediate attractive interactions that promote superconductivity, meaning that the QCP might always be ‘shielded’ by a superconducting dome, just as in Fig. 2. However, there is also typically a diverging correlation length at a QCP, while no such growing correlation length has yet been observed in the strange metal state of the copper oxides for any of the orders that are considered likely candidates. Moreover, according to the marginal Fermi liquid phenomenology⁵, what is needed is a special sort of quantum criticality that is local in space, and so featureless in k .

Is there a QCP involving the termination of pseudogap order inside the superconducting dome? There is evidence for the termination of pseudogap order in a QCP from early specific heat data¹⁰⁸ and from a dispersion anomaly seen in photoemission¹⁰⁹ as well as a vanishing of the Kerr rotation¹⁰² and charge order⁹², with the latest being a divergence in the effective mass seen in quantum oscillation studies¹¹⁰. But which order parameter rules the quantum critical regime¹¹¹, and is that regime large enough to encompass the entire strange metal region? We argued above that the pseudogap is characterized by several competing ordering tendencies. Even more seriously, this quantum critical description should break down at higher temperatures. But the strange metal persists to the highest attainable temperatures.

For a highly correlated fluid, the interactions are large and so probably cannot be treated using any fundamentally perturbative approach which starts with a free particle description. There is a well-developed and extremely successful theoretical solution of this problem applicable to one-dimensional and quasi-one-dimensional electron fluids based on ‘bosonization’, but no such approach exists in higher dimensions. In this context, it is important to seek new approaches—theories that honestly treat the strong correlation physics—even if the connection to the relevant microscopic physics is unclear. This is where the mathematics of string

theory may help: with the so-called holographic duality, one can address the physics of strongly interacting finite density systems. Its central point, the mathematical anti-deSitter to conformal field theory (AdS/CFT) correspondence, has become a focus for modern string theoretical research. Discovered in 1997¹¹², it demonstrates that the two grand theories of physics, which seem unrelated (general relativity and quantum field theory) can become under certain conditions two sides of the same coin. According to the correspondence, there is a ‘holographic’ relation in the sense that the quantum field theory is like a two-dimensional photographic plate with interference fringes encoding the gravitational physics in three dimensions. Most importantly, the difficult-to-solve strong-coupling quantum field theory is mapped to its more easily solved holographic dual, the weak-coupling gravity theory.

Since 2007, the properties of matter at finite density have been the central focus of this ‘holography’ research¹¹³. At low temperatures one finds superconductors, stripe and current phases, and even Fermi liquids. The observable responses of these states are often similar to experimental observations. However, the great difference is in the nature of the strange metal at higher temperatures. The gravity dual tells us that these systems at finite fermion density should form metallic quantum critical phases, where the scale invariance emerges without fine tuning to any special QCP. However, these ‘conformal’ metals (which exhibit Planckian dissipation) are intrinsically unstable, and upon cooling spawn an extensive manifold of stable states. They also have special scaling properties that are different from any conventional quantum critical state¹¹⁴.

Owing to the limitations of the mathematics, holography can only be proved for certain field theories that have no resemblance whatsoever with the electrons in the copper oxides. It is not currently known whether the traits discussed above are ubiquitous emergent phenomena or somehow tied to these special cases. At the least, however, holography may supply powerful new metaphors, teaching physicists to think differently, and leading to new experimental questions.

The overdoped regime

As the doping is increased beyond the doping optimal for the superconductivity, it appears that a real Fermi liquid begins to be established: quantum oscillations indicate a well-developed large Fermi surface, consistent in detail with the prediction of one electron band theory¹¹⁵. This is supported by ARPES measurements where now sharp peaks are observed near this Fermi surface throughout the Brillouin zone (including the antinodes)^{45,116}. Inelastic neutron scattering data indicate a dramatic suppression of magnetic spectral weight near the antiferromagnetic wave vectors, which may be interpreted as a disappearance of the spin-fluctuation pairing glue, explaining why T_c goes down¹¹⁷. On the other hand, recent resonant inelastic X-ray scattering data have demonstrated pronounced spin fluctuations at smaller wave vectors, implying that strong electron correlations persist even in highly overdoped copper oxides^{33,118}. A big question is how different the Fermi liquid at lower temperatures really is from the anomalous strange metal at higher temperatures. ARPES shows that there is only a weak crossover line that separates these two regimes¹¹⁹.

Outlook

Originally inspired by the desire to find out why superconductivity can happen at a high temperature, condensed matter scientists engaged in a relentless effort to unravel the physics of copper oxides. As we have emphasized, there is still plenty of work to do, especially with regards to the physics of competing order in the underdoped regime. The bottom line is that the existing theoretical machinery seems inadequate to describe both the rich physics of the pseudogap phase and the nature of the strange metal phase.

Experimental techniques with which to control correlated electrons are evolving rapidly. Recent examples include the development of atomically precise layer-deposition methods that allow the tailoring of lattice structures¹²⁰ and coherent optical control techniques⁷⁴. In another development, the practitioners of quantum information and string theory have landed in the same territory, finding to their surprise that they are struggling with

many of the same issues as condensed matter physicists. This is also reflected in the fact that some of the key underlying physics has been captured by advanced numerical techniques like the density matrix renormalization group and its descendants²⁹, which were also motivated by quantum information theory. The jury is still out on whether this is a coincidence or signals the onset of a revolution in physics.

Received 28 October; accepted 22 December 2014.

1. Bednorz, J. G. & Müller, K. A. Possible high T_c superconductivity in the Ba-La-Cu-O system. *Z. Phys. B* **64**, 189–193 (1986).
2. Bardeen, J., Cooper, L. N. & Schrieffer, J. R. Theory of superconductivity. *Phys. Rev.* **108**, 1175–1204 (1957).
3. Cohen, M. & Anderson, P. W. Comments on the maximum superconducting transition temperature. In *Superconductivity in d- and f-band Metals* (ed. Douglas, D. H.) 17–27 (American Institute of Physics, 1972).
4. Nagamatsu, J., Nakagawa, N., Muranaka, T., Zenitani, Y. & Akimitsu, J. Superconductivity at 39 K in magnesium diboride. *Nature* **410**, 63–64 (2001).
5. Varma, C. M., Littlewood, P. B., Schmitt-Rink, S., Abrahams, E. & Ruckenstein, A. E. Phenomenology of the normal state of Cu-O high-temperature superconductors. *Phys. Rev. Lett.* **63**, 1996–1999 (1989).
6. Hussey, N. E., Takenaka, K. & Takagi, H. Universality of the Mott-Ioffe-Regel limit of metals. *Phil. Mag.* **84**, 2847–2864 (2004).
7. Fradkin, E., Kivelson, S. A. & Tranquada, J. M. Theory of intertwined orders in high temperature superconductors. Preprint at <http://arXiv.org/abs/1407.4480> (2014).
8. Wollman, D. A., Van Harlingen, D. J., Lee, W. C., Ginsberg, D. M. & Leggett, A. J. Experimental determination of the superconducting pairing state in YBCO from the phase coherence of YBCO-Pb dc SQUIDS. *Phys. Rev. Lett.* **71**, 2134–2137 (1993).
9. Tsuei, C. C. & Kirtley, J. R. Pairing symmetry in cuprate superconductors. *Rev. Mod. Phys.* **72**, 969–1016 (2000).
10. Scalapino, D. J., Loh, E. & Hirsch, J. E. d-wave pairing near a spin-density-wave instability. *Phys. Rev. B* **34**, 8190–8192 (1986).
11. Miyake, K., Schmitt-Rink, S. & Varma, C. M. Spin-fluctuation-mediated even-parity pairing in heavy-fermion superconductors. *Phys. Rev. B* **34**, 6554–6556 (1986).
12. Béal-Monod, M. T., Bourbonnais, C. & Emery, V. J. Possible superconductivity in nearly antiferromagnetic itinerant fermion systems. *Phys. Rev. B* **34**, 7716–7720 (1986).
13. Dahm, T. *et al.* Strength of the spin-fluctuation-mediated pairing interaction in a high-temperature superconductor. *Nature Phys.* **5**, 217–221 (2009).
14. Zaanen, J. Watching rush hour in the world of electrons. *Science* **315**, 1372–1373 (2007).
15. Anderson, P. W. The resonating valence bond state in La_2CuO_4 and superconductivity. *Science* **235**, 1196–1198 (1987).
16. Vaknin, D. *et al.* Antiferromagnetism in $\text{La}_2\text{CuO}_{4-y}$. *Phys. Rev. Lett.* **58**, 2802–2805 (1987).
17. Chakravarty, S., Halperin, B. I. & Nelson, D. R. Low-temperature behavior of two-dimensional quantum antiferromagnets. *Phys. Rev. Lett.* **60**, 1057–1060 (1988).
18. Doiron-Leyraud, N. *et al.* Quantum oscillations and the Fermi surface in an underdoped high- T_c superconductor. *Nature* **447**, 565–568 (2007).
19. Sebastian, S. E. *et al.* Normal-state nodal electronic structure in underdoped high- T_c copper oxides. *Nature* **511**, 61–64 (2014).
20. Zaanen, J. & Gunnarsson, O. Charged magnetic domain lines and the magnetism of high- T_c oxides. *Phys. Rev. B* **40**, 7391–7394 (1989).
21. Emery, V. J. & Kivelson, S. A. Frustrated electronic phase separation and high-temperature superconductors. *Physica C* **209**, 597–621 (1993).
22. Tranquada, J. M., Sternlieb, B. J., Axe, J. D., Nakamura, Y. & Uchida, S. Evidence for stripe correlations of spins and holes in copper oxide superconductors. *Nature* **375**, 561–563 (1995).
23. Raghu, S., Kivelson, S. A. & Scalapino, D. J. Superconductivity in the repulsive Hubbard model: an asymptotically exact weak-coupling solution. *Phys. Rev. B* **81**, 224505 (2010).
24. Scalapino, D. J. A common thread: the pairing interaction for unconventional superconductors. *Rev. Mod. Phys.* **84**, 1383–1417 (2012).
25. Liu, L., Yao, H., Berg, E., White, S. R. & Kivelson, S. A. Phases of the infinite U Hubbard model on square lattices. *Phys. Rev. Lett.* **108**, 126406 (2012).
26. Lee, P. A., Nagaosa, N. & Wen, X.-G. Doping a Mott insulator: physics of high-temperature superconductivity. *Rev. Mod. Phys.* **78**, 17–85 (2006).
27. Paramakanti, A., Randeria, M. & Trivedi, N. High- T_c superconductors: a variational theory of the superconducting state. *Phys. Rev. B* **70**, 054504 (2004).
28. Georges, A., Kotliar, G., Krauth, W. & Rozenberg, M. J. Dynamical mean-field theory of strongly correlated fermion systems and the limit of infinite dimensions. *Rev. Mod. Phys.* **68**, 13–125 (1996).
29. Stoudenmire, E. M. & White, S. R. Studying two-dimensional systems with the density matrix renormalization group. *Annu. Rev. Condens. Matter Phys.* **3**, 61–6.18 (2012).
30. White, S. R. & Scalapino, D. J. Density matrix renormalization group study of the striped phase in the 2D t-J model. *Phys. Rev. Lett.* **80**, 1272–1275 (1998).
31. Corboz, P., Rice, T. M. & Troyer, M. Competing states in the t-J model: uniform d-wave state versus stripe state. *Phys. Rev. Lett.* **113**, 046402 (2014).
32. Le Tacon, M. *et al.* Intense paramagnon excitations in a large family of high-temperature superconductors. *Nature Phys.* **7**, 725–730 (2011).
33. Dean, M. P. M. *et al.* Persistence of magnetic excitations in $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$ from the undoped insulator to the heavily overdoped non-superconducting metal. *Nature Mater.* **12**, 1019–1023 (2013).
34. Fong, H. F. *et al.* Spin susceptibility in underdoped $\text{YBa}_2\text{Cu}_3\text{O}_{6+x}$. *Phys. Rev. B* **61**, 14773–14786 (2000).
35. Dai, P., Mook, H. A., Hunt, R. D. & Dogan, F. Evolution of the resonance and incommensurate spin fluctuations in superconducting $\text{YBa}_2\text{Cu}_3\text{O}_{6+x}$. *Phys. Rev. B* **63**, 054525 (2001).
36. Carbotte, J. P., Timusk, T. & Hwang, J. Bosons in high-temperature superconductors: an experimental survey. *Rep. Prog. Phys.* **74**, 066501 (2011).
37. Anderson, P. W. Is there glue in cuprate superconductors? *Science* **316**, 1705–1707 (2007).
38. Lanzara, A. *et al.* Evidence for ubiquitous strong electron-phonon coupling in high-temperature superconductors. *Nature* **412**, 510–514 (2001).
39. Reznik, D. *et al.* Electron-phonon coupling reflecting dynamic charge inhomogeneity in copper oxide superconductors. *Nature* **440**, 1170–1173 (2006).
40. Leggett, A. J. A “midinfrared” scenario for cuprate superconductivity. *Proc. Natl Acad. Sci. USA* **96**, 8365–8372 (1999).
41. Abrikosov, A. A. & Gor’kov, L. P. Contribution to the theory of superconducting alloys with paramagnetic impurities. *Sov. Phys. JETP* **12**, 1243–1253 (1961).
42. Garg, A., Randeria, M. & Trivedi, N. Strong correlations make high-temperature superconductors robust against disorder. *Nature Phys.* **4**, 762–765 (2008).
43. Uemura, Y. J. *et al.* Universal correlations between T_c and n_0/m^* in high- T_c cuprate superconductors. *Phys. Rev. Lett.* **62**, 2317–2320 (1989).
44. Emery, V. J. & Kivelson, S. A. Importance of phase fluctuations in superconductors with small superfluid density. *Nature* **374**, 434–437 (1995).
45. Chatterjee, U. *et al.* Electronic phase diagram of high-temperature copper oxide superconductors. *Proc. Natl Acad. Sci. USA* **108**, 9346–9349 (2011).
46. Matsui, H. *et al.* BCS-like Bogoliubov quasiparticles in high- T_c superconductors observed by angle-resolved photoemission spectroscopy. *Phys. Rev. Lett.* **90**, 217002 (2003).
47. Fedorov, A. V. *et al.* Temperature dependent photoemission studies of optimally doped $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8-\delta}$. *Phys. Rev. Lett.* **82**, 2179–2182 (1999).
48. Feng, D. L. *et al.* Signature of superfluid density in the single-particle excitation spectrum of $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8-\delta}$. *Science* **289**, 277–281 (2000).
49. Warren, W. W. *et al.* Cu spin dynamics and superconducting precursor effects in planes above T_c in $\text{YBa}_2\text{Cu}_3\text{O}_{6.7}$. *Phys. Rev. Lett.* **62**, 1193–1196 (1989).
50. Alloul, H., Ohno, T. & Mendels, P. ^{89}Y NMR evidence for a Fermi-liquid behavior in $\text{YBa}_2\text{Cu}_3\text{O}_{6+x}$. *Phys. Rev. Lett.* **63**, 1700–1703 (1989).
51. Homes, C. C., Timusk, T., Liang, R., Bonn, D. A. & Hardy, W. N. Optical conductivity of c axis oriented $\text{YBa}_2\text{Cu}_3\text{O}_{6.70}$: evidence for a pseudogap. *Phys. Rev. Lett.* **71**, 1645–1648 (1993).
52. Puchkov, A. V., Basov, D. N. & Timusk, T. The pseudogap state in high- T_c superconductors: an infrared study. *J. Phys. Condens. Matter* **8**, 10049–10082 (1996).
53. Bucher, B., Steiner, P., Karpinski, J., Kaldis, E. & Wachter, P. Influence of the spin gap on the normal state transport in $\text{YBa}_2\text{Cu}_4\text{O}_8$. *Phys. Rev. Lett.* **70**, 2012–2015 (1993).
54. Ito, T., Takenaka, K. & Uchida, S. Systematic deviation from T-linear behavior in the in-plane resistivity of $\text{YBa}_2\text{Cu}_3\text{O}_{7-x}$: evidence for dominant spin scattering. *Phys. Rev. Lett.* **70**, 3995–3998 (1993).
55. Hashimoto, M. *et al.* Energy gaps in high-transition-temperature cuprate superconductors. *Nature Phys.* **10**, 483–495 (2014).
56. Norman, M. R. *et al.* Destruction of the Fermi surface in underdoped high- T_c superconductors. *Nature* **392**, 157–160 (1998).
57. Marshall, D. S. *et al.* Unconventional electronic structure evolution with hole doping in $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$: angle-resolved photoemission results. *Phys. Rev. Lett.* **76**, 4841–4844 (1996).
58. Yang, H.-B. *et al.* Emergence of preformed Cooper pairs from the doped Mott insulating state in $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$. *Nature* **456**, 77–80 (2008).
59. Hoffman, J. E. *et al.* Imaging quasiparticle interference in $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$. *Science* **297**, 1148–1151 (2002).
60. McElroy, K. *et al.* Relating atomic-scale electronic phenomena to wave-like quasiparticle states in superconducting $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$. *Nature* **422**, 592–596 (2003).
61. Kohsaka, Y. *et al.* How Cooper pairs vanish approaching the Mott insulator in $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$. *Nature* **454**, 1072–1078 (2008).
62. Fujita, K. *et al.* Simultaneous transitions in cuprate momentum-space topology and electronic symmetry breaking. *Science* **344**, 612–616 (2014).

63. Vishik, I. M. *et al.* A momentum-dependent perspective on quasiparticle interference in $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$. *Nature Phys.* **5**, 718–721 (2009).
64. Vershinin, M. *et al.* Local ordering in the pseudogap state of the high- T_c superconductor $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$. *Science* **303**, 1995–1998 (2004).
65. Parker, C. V. *et al.* Fluctuating stripes at the onset of the pseudogap in the high- T_c superconductor $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$. *Nature* **468**, 677–680 (2010).
66. Howald, C., Eisaki, H., Kaneko, N. & Kapitulnik, A. Coexistence of periodic modulation of quasiparticle states and superconductivity in $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$. *Proc. Natl Acad. Sci. USA* **100**, 9705–9709 (2003).
67. da Silva Neto, E. H. *et al.* Ubiquitous interplay between charge ordering and high-temperature superconductivity in cuprates. *Science* **343**, 393–396 (2014).
68. Comin, R. *et al.* Charge order driven by Fermi-arc instability in $\text{Bi}_2\text{Sr}_{2-x}\text{La}_x\text{CuO}_{6+\delta}$. *Science* **343**, 390–392 (2014).
69. Ding, H. *et al.* Spectroscopic evidence for a pseudogap in the normal state of underdoped high- T_c superconductors. *Nature* **382**, 51–54 (1996).
70. Loeser, A. G. *et al.* Excitation gap in the normal state of underdoped $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$. *Science* **273**, 325–329 (1996).
71. Corson, J., Mallozzi, R., Orenstein, J., Eckstein, J. N. & Bozovic, I. Vanishing of phase coherence in underdoped $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$. *Nature* **398**, 221–223 (1999).
72. Li, L. *et al.* Diamagnetism and Cooper pairing above T_c in cuprates. *Phys. Rev. B* **81**, 054510 (2010).
73. Dubroka, A. *et al.* Evidence of a precursor superconducting phase at temperatures as high as 180 K in $\text{RBa}_2\text{Cu}_3\text{O}_{7-\delta}$ ($R = \text{Y, Gd, Eu}$) superconducting crystals from infrared spectroscopy. *Phys. Rev. Lett.* **106**, 047006 (2011).
74. Kaiser, S. *et al.* Optically induced coherent transport far above T_c in underdoped $\text{YBa}_2\text{Cu}_3\text{O}_{6+\delta}$. *Phys. Rev. B* **89**, 184516 (2014).
75. Norman, M. R. *et al.* Modeling the Fermi arc in underdoped cuprates. *Phys. Rev. B* **76**, 174501 (2007).
76. Reber, T. J. *et al.* The origin and non-quasiparticle nature of Fermi arcs in $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$. *Nature Phys.* **8**, 606–610 (2012).
77. Lee, J. *et al.* Spectroscopic fingerprint of phase-incoherent superconductivity in the underdoped $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$. *Science* **325**, 1099–1103 (2009).
78. Chakravarty, S., Laughlin, R. B., Morr, D. K. & Nayak, C. Hidden order in the cuprates. *Phys. Rev. B* **63**, 094503 (2001).
79. Scalapino, D. J. & White, S. R. Stripe structures in the t-t'-J model. *Physica C* **481**, 146–152 (2012).
80. Tajima, S., Noda, T., Eisaki, H. & Uchida, S. c-Axis optical response in the static stripe ordered phase of the cuprates. *Phys. Rev. Lett.* **86**, 500–503 (2001).
81. Li, Q., Hücker, M., Gu, G. D., Tsvetlik, A. M. & Tranquada, J. M. Two-dimensional superconducting fluctuations in stripe-ordered $\text{La}_{1.875}\text{Ba}_{0.125}\text{CuO}_4$. *Phys. Rev. Lett.* **99**, 067001 (2007).
82. Arai, M. *et al.* Incommensurate spin dynamics of underdoped superconductor $\text{YBa}_2\text{Cu}_3\text{O}_{6.7}$. *Phys. Rev. Lett.* **83**, 608–611 (1999).
83. Ulbrich, H. & Braden, M. Neutron scattering studies on stripe phases in non-cuprate materials. *Physica C* **481**, 31–45 (2012).
84. Hinkov, V. *et al.* Spin dynamics in the pseudogap state of a high-temperature superconductor. *Nature Phys.* **3**, 780–785 (2007).
85. Norman, M. R. Linear response theory and the universal nature of the magnetic excitation spectrum of the cuprates. *Phys. Rev. B* **75**, 184514 (2007).
86. Wu, T. *et al.* Magnetic-field-induced charge-stripe order in the high-temperature superconductor $\text{YBa}_2\text{Cu}_3\text{O}_y$. *Nature* **477**, 191–194 (2011).
87. Tabis, W. *et al.* Connection between charge-density-wave order and charge transport in the cuprate superconductors. *Nature Commun.* **5**, 5875 (2014).
88. Ghiringhelli, G. *et al.* Long-range incommensurate charge fluctuations in $(\text{Y,Nd})\text{Ba}_2\text{Cu}_3\text{O}_{6+x}$. *Science* **337**, 821–825 (2012).
- Demonstration of charge crystallization without spin order as a generic ordering phenomenon in underdoped copper oxides, using resonant X-ray scattering.**
89. Chang, J. *et al.* Direct observation of competition between superconductivity and charge density wave order in $\text{YBa}_2\text{Cu}_3\text{O}_{6.67}$. *Nature Phys.* **8**, 871–876 (2012).
90. Le Tacon, M. *et al.* Inelastic X-ray scattering in $\text{YBa}_2\text{Cu}_3\text{O}_{6.6}$ reveals giant phonon anomalies and elastic central peak due to charge-density-wave formation. *Nature Phys.* **10**, 52–58 (2014).
91. Wu, T. *et al.* Short-range charge order reveals the role of disorder in the pseudogap state of high- T_c superconductors. Preprint at <http://arXiv.org/abs/1404.1617> (2014).
92. Blanco-Canosa, S. *et al.* Resonant X-ray scattering study of charge density wave correlations in $\text{YBa}_2\text{Cu}_3\text{O}_{6+x}$. *Phys. Rev. B* **90**, 054513 (2014).
93. Kivelson, S. A., Fradkin, E. & Emery, V. J. Electronic liquid-crystal phases of a doped Mott insulator. *Nature* **393**, 550–553 (1998).
94. Ando, Y., Segawa, K., Komiya, S. & Lavrov, A. N. Electrical resistivity anisotropy from self-organized one dimensionality in high-temperature superconductors. *Phys. Rev. Lett.* **88**, 137005 (2002).
95. Daou, R. *et al.* Broken rotational symmetry in the pseudogap phase of a high- T_c superconductor. *Nature* **463**, 519–522 (2010).
96. Hinkov, V. *et al.* Electronic liquid crystal state in the high-temperature superconductor $\text{YBa}_2\text{Cu}_3\text{O}_{6.45}$. *Science* **319**, 597–600 (2008).
97. Kivelson, S. A. *et al.* How to detect fluctuating stripes in the high-temperature superconductors. *Rev. Mod. Phys.* **75**, 1201–1241 (2003).
- A pedagogical treatise of electronic charge order and how to study it experimentally.**
98. Lawler, M. J. *et al.* Intra-unit-cell electronic nematicity of the high- T_c copper-oxide pseudogap states. *Nature* **466**, 347–351 (2010).
99. da Silva Neto, E. H. *et al.* Detection of electronic nematicity using scanning tunnelling microscopy. *Phys. Rev. B* **87**, 161117 (2013).
100. Fauqué, B. *et al.* Magnetic order in the pseudogap phase of high- T_c superconductors. *Phys. Rev. Lett.* **96**, 197001 (2006).
101. Varma, C. M. Non-Fermi-liquid states and pairing instability of a general model of copper oxide metals. *Phys. Rev. B* **55**, 14554–14580 (1997).
102. Xia, J. *et al.* Polar Kerr-effect measurements of the high-temperature $\text{YBa}_2\text{Cu}_3\text{O}_{6+x}$ superconductor: evidence for broken symmetry near the pseudogap temperature. *Phys. Rev. Lett.* **100**, 127002 (2008).
103. Martin, S., Fiory, A. T., Fleming, R. M., Schneemeyer, L. F. & Waszczak, J. V. Normal-state transport properties of $\text{Bi}_{2-x}\text{Sr}_x\text{CuO}_{6+\delta}$ crystals. *Phys. Rev. B* **41**, 846–849 (1990).
104. Chien, T. R., Wang, Z. Z. & Ong, N. P. Effect of Zn impurities on the normal-state Hall angle in single-crystal $\text{YBa}_2\text{Cu}_3\text{Zn}_{0.7}\text{O}_{7-\delta}$. *Phys. Rev. Lett.* **67**, 2088–2091 (1991).
105. Valla, T. *et al.* Temperature dependent scattering rates at the Fermi surface of optimally doped $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$. *Phys. Rev. Lett.* **85**, 828–831 (2000).
106. Sachdev, S. *Quantum Phase Transitions* (Cambridge Univ. Press, 1999).
- An authoritative text covering the fundamentals of quantum phase transitions.**
107. Zaanen, J. Superconductivity: why the temperature is high. *Nature* **430**, 512–513 (2004).
108. Tallon, J. L., Williams, G. V. M., Staines, M. P. & Bernhard, C. Energy and length scales in the superconducting phase diagram for HTSC cuprates. *Physica C* **235–240**, 1821–1822 (1994).
109. Vishik, I. M. *et al.* Phase competition in trisected superconducting dome. *Proc. Natl Acad. Sci. USA* **109**, 18332–18337 (2012).
110. Ramshaw, B. J. *et al.* A quantum critical point at the heart of high temperature superconductivity. Preprint at <http://arXiv.org/abs/1409.3990> (2014).
111. Castellani, C., Di Castro, C. & Grilli, M. Singular quasiparticle scattering in the proximity of charge instabilities. *Phys. Rev. Lett.* **75**, 4650–4653 (1995).
112. Maldacena, J. M. The large N limit of superconformal field theories and supergravity. *Adv. Theor. Math. Phys.* **2**, 231–252 (1998).
113. Zaanen, J., Sun, Y. W., Liu, Y. & Schalm, K. *Holographic Duality for Condensed Matter Physics* (Cambridge Univ. Press, in the press).
- A comprehensive but accessible text for condensed matter applications of the AdS/CFT correspondence.**
114. Iqbal, N., Liu, H. & Mezei, M. Lectures on holographic non-Fermi liquids and quantum phase transitions. In *String Theory and Its Applications, TASI 2010*, (eds Dine, M., Banks, T. & Sachdev, S.) 707–816 (World Scientific, 2012).
115. Vignolle, D. *et al.* Quantum oscillations in an overdoped high- T_c superconductor. *Nature* **455**, 952–955 (2008).
116. Platé, M. *et al.* Fermi surface and quasiparticle excitations of overdoped $\text{Ti}_2\text{Ba}_2\text{CuO}_{6+\delta}$. *Phys. Rev. Lett.* **95**, 077001 (2005).
117. Wakimoto, S. *et al.* Direct relation between the low-energy spin excitations and superconductivity of overdoped high- T_c superconductors. *Phys. Rev. Lett.* **92**, 217004 (2004).
118. Le Tacon, M. *et al.* Dispersive spin excitations in highly overdoped cuprates revealed by resonant inelastic x-ray scattering. *Phys. Rev. B* **88**, 020501 (2013).
119. Kaminski, A. *et al.* Crossover from coherent to incoherent electronic excitations in the normal state of $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$. *Phys. Rev. Lett.* **90**, 207003 (2003).
120. Gozar, A. *et al.* High-temperature interface superconductivity between metallic and insulating copper oxides. *Nature* **455**, 782–785 (2008).

Acknowledgements We thank A. Yazdani for many discussions. S.A.K. was supported by the US DOE, Basic Energy Sciences, Materials Science and Engineering, under Award No. DE-AC02-76SF00515 at Stanford University. M.N. was supported by the Center for Emergent Superconductivity, an Energy Frontier Research Center funded by the US DOE, Basic Energy Sciences, under Award No. DE-AC0298CH1088. J.Z. acknowledges financial support by the Netherlands Organization for Scientific Research/Ministry of Science and Education (NWO/OCW), and a grant from the Templeton foundation: the opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the John Templeton foundation.

Author Contributions All authors contributed to the text. S.U. prepared the figures.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.Z. (jan@lorenz.leidenuniv.nl).

New genetic loci link adipose and insulin biology to body fat distribution

A list of authors and their affiliations appears at the end of the paper

Body fat distribution is a heritable trait and a well-established predictor of adverse metabolic outcomes, independent of overall adiposity. To increase our understanding of the genetic basis of body fat distribution and its molecular links to cardiometabolic traits, here we conduct genome-wide association meta-analyses of traits related to waist and hip circumferences in up to 224,459 individuals. We identify 49 loci (33 new) associated with waist-to-hip ratio adjusted for body mass index (BMI), and an additional 19 loci newly associated with related waist and hip circumference measures ($P < 5 \times 10^{-8}$). In total, 20 of the 49 waist-to-hip ratio adjusted for BMI loci show significant sexual dimorphism, 19 of which display a stronger effect in women. The identified loci were enriched for genes expressed in adipose tissue and for putative regulatory elements in adipocytes. Pathway analyses implicated adipogenesis, angiogenesis, transcriptional regulation and insulin resistance as processes affecting fat distribution, providing insight into potential pathophysiological mechanisms.

Depot-specific accumulation of fat, particularly in the central abdomen, confers an increased risk of metabolic and cardiovascular diseases and mortality¹. An easily accessible measure of body fat distribution is waist-to-hip ratio (WHR), a comparison of waist and hip circumferences. A larger WHR indicates more intra-abdominal fat deposition and is associated with higher risk for type 2 diabetes (T2D) and cardiovascular disease^{2,3}. Conversely, a smaller WHR indicates greater gluteal fat accumulation and is associated with lower risk for T2D, hypertension, dyslipidemia and mortality^{4–6}. Our previous genome-wide association study (GWAS) meta-analyses have identified loci for WHR after adjusting for body mass index (WHRadjBMI)^{7,8}. These loci are enriched for association with other metabolic traits^{7,8} and show that different fat distribution patterns can have distinct genetic components^{9,10}.

To determine further the genetic architecture of fat distribution and to increase our understanding of molecular connections with cardiometabolic traits, we performed a meta-analysis of WHRadjBMI associations in 142,762 individuals with GWAS data and 81,697 individuals genotyped with the Metabochip¹¹, all from the Genetic Investigation of ANthropometric Traits (GIANT) consortium. Given the marked sexual dimorphism previously observed among established WHRadjBMI loci^{7,8}, we performed analyses in men and women separately, the results of which were subsequently combined. To characterize the genetic determinants of specific aspects of body fat distribution more fully, we performed secondary GWAS meta-analyses for five additional traits: unadjusted WHR, unadjusted waist circumference, BMI-adjusted waist circumference (WCadjBMI), unadjusted hip circumference and BMI-adjusted hip circumference (HIPadjBMI). We evaluated the associated loci to understand their contributions to variation in fat distribution and adipose tissue biology, and their molecular links to cardiometabolic traits.

New loci associated with WHRadjBMI

We performed meta-analyses of GWAS of WHRadjBMI in up to 142,762 individuals of European ancestry from 57 new or previously described GWAS⁷, and separately in up to an additional 67,326 European ancestry individuals from 44 Metabochip studies (Extended Data Fig. 1 and Supplementary Tables 1–3). The combination of these two meta-analyses included up to 2,542,447 autosomal single nucleotide polymorphisms (SNPs) in up to 210,088 European ancestry individuals. We defined new loci based on genome-wide significant association

($P < 5 \times 10^{-8}$ after genomic control correction at both the study-specific and meta-analytic levels) and distance (>500 kilobases (kb) from previously established loci)^{7,8}.

We identified 49 loci for WHRadjBMI, 33 of which were new and 16 previously described^{7,8}. Of these, a European ancestry sex-combined analysis identified 39 loci, 24 of which were new^{7,8} (Table 1, Supplementary Table 4 and Supplementary Figs 1–3). European ancestry sex-specific analyses identified nine additional loci, eight of which were new and significant in women but not in men (all $P_{\text{men}} > 0.05$; Table 1 and Supplementary Fig. 4). The addition of 14,371 individuals of non-European ancestry genotyped on the Metabochip identified one additional locus in women (rs1534696, near *SNX10*, $P_{\text{women}} = 2.1 \times 10^{-8}$, $P_{\text{men}} = 0.26$, Table 1 and Supplementary Tables 1–3), with no evidence of heterogeneity across ancestries ($P_{\text{het}} = 0.86$; Supplementary Note).

Genetic architecture of WHRadjBMI

To evaluate sexual dimorphism, we compared sex-specific effect size estimates of the 49 WHRadjBMI lead SNPs. The effect estimates were significantly different ($P_{\text{difference}} < 0.05/49 = 0.001$) at 20 SNPs, 19 of which showed larger effects in women (Table 1 and Extended Data Fig. 2a), similar to previous findings^{7,8}. The only SNP that showed a larger effect in men mapped near *GDF5* (rs224333, $\beta_{\text{men}} = 0.036$ and $P = 9.0 \times 10^{-12}$, $\beta_{\text{women}} = 0.009$ and $P = 0.074$, $P_{\text{difference}} = 6.4 \times 10^{-5}$), a locus previously associated with height (rs6060369, $r^2 = 0.96$ and rs143384, $r^2 = 0.96$, 1000 Genomes Project CEU), although without significant differences between sexes^{12,13}. Consistent with the larger number of loci identified in women, variance component analyses demonstrated a significantly larger heritability (h^2) of WHRadjBMI in women than men in the Framingham Heart Study ($h^2_{\text{women}} = 0.46$, $h^2_{\text{men}} = 0.19$, $P_{\text{difference}} = 0.0037$) and TwinGene study ($h^2_{\text{women}} = 0.56$, $h^2_{\text{men}} = 0.32$, $P_{\text{difference}} = 0.001$; Supplementary Table 5 and Extended Data Fig. 2b).

To identify multiple association signals within observed loci, we performed approximate conditional analyses of the sex-combined and sex-specific summary statistics using GCTA¹⁴ (Supplementary Note). Several signals ($P < 5 \times 10^{-8}$) were identified at nine loci (Extended Data Table 1). Fitting SNPs jointly identified different lead SNPs in the sex-specific and sex-combined analyses. For example, the *MAP3K1-ANKRD55* locus showed near-independent (linkage disequilibrium (LD) $r^2 < 0.01$) SNPs 54 kb apart that were significant only in women

Table 1 | WHRadjBMI loci in sex-combined and sex-specific meta-analyses

SNP	Chr	Locus	EA*	EAF	Sex-combined			Women			Men			Sex diff.
					β	P	N	β	P	N	β	P	N	
Novel loci achieving genome-wide significance in European-ancestry meta-analyses														
rs905938	1	DCST2	T	0.74	0.025	7.3×10 ⁻¹⁰	207,867	0.034	4.9×10 ⁻¹⁰	115,536	0.015	1.1×10 ⁻²	92,461	1.6×10 ⁻²
rs10919388	1	GORAB	C	0.72	0.024	3.2×10 ⁻⁹	181,049	0.033	4.8×10 ⁻¹⁰	102,446	0.013	3.0×10 ⁻²	78,738	9.8×10 ⁻³
rs1385167	2	MEIS1	G	0.15	0.029	1.9×10 ⁻⁹	206,619	0.023	4.0×10 ⁻⁴	114,668	0.036	2.3×10 ⁻⁷	92,085	1.6×10 ⁻¹
rs1569135	2	CALCR1	A	0.53	0.021	5.6×10 ⁻¹⁰	209,906	0.023	6.9×10 ⁻⁷	116,642	0.019	1.5×10 ⁻⁴	93,398	5.8×10 ⁻¹
rs10804591	3	PLXND1	A	0.79	0.025	6.6×10 ⁻⁹	209,921	0.040	6.1×10 ⁻¹³	116,667	0.004	5.3×10 ⁻¹	93,387	5.7×10 ⁻⁶
rs17451107	3	LEKR1	T	0.61	0.026	1.1×10 ⁻¹²	207,795	0.023	1.0×10 ⁻⁶	115,735	0.030	1.4×10 ⁻⁸	92,194	3.5×10 ⁻¹
rs3805389	4	NMU	A	0.28	0.012	1.5×10 ⁻³	209,218	0.027	4.6×10 ⁻⁸	116,226	-0.007	2.1×10 ⁻¹	93,125	1.6×10 ⁻⁶
rs9991328	4	FAM13A	T	0.49	0.019	4.5×10 ⁻⁸	209,925	0.028	3.4×10 ⁻¹⁰	116,652	0.007	1.7×10 ⁻¹	93,407	8.5×10 ⁻⁴
rs303084	4	SPATA5-FGF2	A	0.80	0.023	3.9×10 ⁻⁸	209,941	0.029	3.4×10 ⁻⁷	116,662	0.016	9.9×10 ⁻³	93,412	1.1×10 ⁻¹
rs9687846	5	MAP3K1	A	0.19	0.024	7.1×10 ⁻⁸	208,181	0.041	3.8×10 ⁻¹²	115,897	0.000	9.7×10 ⁻¹	92,417	1.3×10 ⁻⁶
rs6556301	5	FGFR4	T	0.36	0.022	2.6×10 ⁻⁸	178,874	0.018	7.1×10 ⁻⁴	101,638	0.029	1.0×10 ⁻⁶	77,370	1.4×10 ⁻¹
rs7759742	6	BTNL2	A	0.51	0.023	4.4×10 ⁻¹¹	208,263	0.024	1.7×10 ⁻⁷	115,648	0.023	5.5×10 ⁻⁶	92,749	8.6×10 ⁻¹
rs1776897	6	HMGAI1	G	0.08	0.030	1.1×10 ⁻⁵	177,879	0.052	6.8×10 ⁻⁹	100,516	0.003	7.4×10 ⁻¹	77,497	1.8×10 ⁻⁴
rs7801581	7	HXA11	T	0.24	0.027	3.7×10 ⁻¹⁰	195,215	0.025	7.7×10 ⁻⁶	108,866	0.029	2.4×10 ⁻⁶	86,483	6.9×10 ⁻¹
rs7830933	8	NKX2-6	A	0.77	0.022	7.4×10 ⁻⁸	209,766	0.037	1.2×10 ⁻¹²	116,567	0.001	8.4×10 ⁻¹	93,333	1.4×10 ⁻⁶
rs12679556	8	MSC	G	0.25	0.027	2.1×10 ⁻¹¹	203,826	0.033	2.1×10 ⁻¹⁰	114,369	0.017	4.2×10 ⁻³	89,591	2.8×10 ⁻²
rs10991437	9	ABCA1	A	0.11	0.031	1.0×10 ⁻⁸	209,941	0.040	2.8×10 ⁻⁸	116,644	0.022	6.1×10 ⁻³	93,430	7.2×10 ⁻²
rs7917772	10	SFXN2	A	0.62	0.014	5.6×10 ⁻⁵	209,642	0.027	5.5×10 ⁻⁹	116,514	-0.001	8.6×10 ⁻¹	93,263	2.3×10 ⁻⁵
rs11231693	11	MACROD1-VEGFB	A	0.06	0.041	4.5×10 ⁻⁸	198,072	0.068	2.7×10 ⁻¹¹	110,164	0.009	4.2×10 ⁻¹	88,043	2.5×10 ⁻⁵
rs4765219	12	CCDC92	C	0.67	0.028	1.6×10 ⁻¹⁵	209,807	0.037	1.0×10 ⁻¹⁴	116,592	0.018	5.3×10 ⁻⁴	93,350	5.7×10 ⁻³
rs8042543	15	KLF13	C	0.78	0.026	1.2×10 ⁻⁹	208,255	0.023	6.7×10 ⁻⁵	115,760	0.030	1.0×10 ⁻⁶	92,629	3.6×10 ⁻¹
rs8030605	15	RFX7	A	0.14	0.030	8.8×10 ⁻⁹	208,374	0.031	1.0×10 ⁻⁵	115,864	0.031	5.9×10 ⁻⁵	92,644	9.9×10 ⁻¹
rs1440372	15	SMAD6	C	0.71	0.024	1.1×10 ⁻¹⁰	207,447	0.022	1.1×10 ⁻⁵	115,201	0.027	1.4×10 ⁻⁶	92,380	5.2×10 ⁻¹
rs2925979	16	CMIP	T	0.31	0.018	1.2×10 ⁻⁶	207,828	0.032	3.4×10 ⁻¹¹	115,431	-0.002	7.9×10 ⁻¹	92,531	1.2×10 ⁻⁶
rs4646404	17	PMT	G	0.67	0.027	1.4×10 ⁻¹¹	198,196	0.034	5.3×10 ⁻¹¹	115,337	0.017	2.5×10 ⁻³	87,857	2.6×10 ⁻²
rs8066985	17	KCNJ2	A	0.50	0.018	1.4×10 ⁻⁷	209,977	0.026	4.0×10 ⁻⁹	116,683	0.007	1.9×10 ⁻¹	93,428	1.8×10 ⁻³
rs12454712	18	BCL2	T	0.61	0.016	1.0×10 ⁻⁴	169,793	0.035	1.1×10 ⁻⁹	96,182	-0.007	2.5×10 ⁻¹	73,576	1.6×10 ⁻⁷
rs12608504	19	JUND	A	0.36	0.022	8.8×10 ⁻¹⁰	209,990	0.017	2.6×10 ⁻⁴	116,689	0.028	1.1×10 ⁻⁷	93,435	1.2×10 ⁻¹
rs4081724	19	CEBPA	G	0.85	0.035	7.4×10 ⁻¹²	207,418	0.033	9.2×10 ⁻⁷	115,322	0.039	1.4×10 ⁻⁷	92,230	5.0×10 ⁻¹
rs979012	20	BMP2	T	0.34	0.027	3.3×10 ⁻¹⁴	209,941	0.026	1.0×10 ⁻⁷	116,668	0.028	6.6×10 ⁻⁸	93,407	6.7×10 ⁻¹
rs224333	20	GDF5	G	0.62	0.020	2.6×10 ⁻⁸	208,025	0.009	7.4×10 ⁻²	115,803	0.036	9.0×10 ⁻¹²	92,356	6.4×10 ⁻⁵
rs6090583	20	EYA2	A	0.48	0.022	6.2×10 ⁻¹¹	209,435	0.029	2.8×10 ⁻¹⁰	116,382	0.015	2.4×10 ⁻³	93,187	3.2×10 ⁻²
Novel loci achieving genome-wide significance in all-ancestry meta-analyses														
rs1534696	7	SNX10	C	0.43	0.011	1.3×10 ⁻³	212,501	0.027	2.1×10 ⁻⁸	118,187	-0.006	2.6×10 ⁻¹	92,243	2.1×10 ⁻⁶
Previously reported loci achieving genome-wide significance in European-ancestry meta-analyses														
rs2645294	1	TBX15-WARS2	T	0.58	0.031	1.7×10 ⁻¹⁹	209,808	0.035	1.5×10 ⁻¹⁴	116,596	0.027	1.5×10 ⁻⁷	93,346	2.0×10 ⁻¹
rs714515	1	DNM3-PIGC	G	0.43	0.027	4.4×10 ⁻¹⁵	203,401	0.029	1.8×10 ⁻¹⁰	113,939	0.025	8.5×10 ⁻⁷	89,596	5.1×10 ⁻¹
rs2820443	1	LYPLAL1	T	0.72	0.035	5.3×10 ⁻²¹	209,975	0.062	5.7×10 ⁻³⁵	116,672	0.002	6.9×10 ⁻¹	93,437	2.6×10 ⁻¹⁷
rs10195252	2	GRB14-COBL1	T	0.59	0.027	5.9×10 ⁻¹⁵	209,395	0.052	4.7×10 ⁻³⁰	116,329	-0.003	5.3×10 ⁻¹	93,199	2.4×10 ⁻¹⁷
rs17819328	3	PPARG	G	0.43	0.021	2.4×10 ⁻⁹	208,809	0.035	4.6×10 ⁻¹⁴	116,072	0.005	3.3×10 ⁻¹	92,871	5.1×10 ⁻⁶
rs2276824	3	PBRM1†	C	0.43	0.024	3.2×10 ⁻¹¹	208,901	0.028	3.7×10 ⁻⁹	116,128	0.020	1.4×10 ⁻⁴	92,907	2.0×10 ⁻¹
rs2371767	3	ADAMTS9	G	0.72	0.036	1.6×10 ⁻²⁰	194,506	0.056	1.2×10 ⁻²⁶	108,624	0.012	3.5×10 ⁻²	86,016	3.6×10 ⁻⁹
rs1045241	5	TNFAIP8-HSD17B4	C	0.71	0.019	4.4×10 ⁻⁷	209,710	0.035	6.6×10 ⁻¹²	116,560	-0.001	9.3×10 ⁻¹	93,284	8.3×10 ⁻⁷
rs7705502	5	CPEB4	A	0.33	0.027	4.7×10 ⁻¹⁴	209,827	0.027	1.9×10 ⁻⁸	116,609	0.027	2.3×10 ⁻⁷	93,352	>0.99
rs1294410	6	LY86	C	0.63	0.031	2.0×10 ⁻¹⁸	209,830	0.037	1.6×10 ⁻¹⁵	116,624	0.025	1.4×10 ⁻⁶	93,340	6.3×10 ⁻²
rs1358980	6	VEGFA	T	0.47	0.039	3.1×10 ⁻²⁷	206,862	0.060	3.7×10 ⁻³⁴	115,047	0.015	4.0×10 ⁻³	91,949	3.7×10 ⁻¹¹
rs1936805	6	RSPO3	T	0.51	0.043	3.6×10 ⁻³⁵	209,859	0.052	3.7×10 ⁻³⁰	116,602	0.031	3.1×10 ⁻¹⁰	93,392	1.0×10 ⁻³
rs10245353	7	NFE2L3	A	0.20	0.035	8.4×10 ⁻¹⁶	210,008	0.041	7.9×10 ⁻¹³	116,704	0.027	1.4×10 ⁻⁵	93,438	7.2×10 ⁻²
rs10842707	12	ITPR2-SSPN	T	0.23	0.032	4.4×10 ⁻¹⁶	210,023	0.041	6.1×10 ⁻¹⁵	116,704	0.022	1.4×10 ⁻⁴	93,453	1.1×10 ⁻²
rs1443512	12	HOXC13	A	0.24	0.028	6.9×10 ⁻¹³	209,980	0.040	1.1×10 ⁻¹⁴	116,688	0.013	2.8×10 ⁻²	93,425	1.6×10 ⁻⁴
rs2294239	22	ZNRF3	A	0.59	0.025	7.2×10 ⁻¹³	209,454	0.028	6.9×10 ⁻¹⁰	116,414	0.024	2.3×10 ⁻⁶	93,173	5.0×10 ⁻¹

Loci achieving genome-wide significance ($P < 5 \times 10^{-8}$) in sex-combined and/or sex-specific meta-analyses. *P* values and β coefficients for the association with WHRadjBMI in the meta-analyses of combined GWAS and MetaboChip studies. The smallest *P* value for each SNP is shown in bold. Chr, chromosome; EAF, effect allele frequency.

* The effect allele (EA) is the WHRadjBMI-increasing allele in the sex-combined analysis.

† Test for sex difference; values significant at the table-wide Bonferroni threshold of $0.05/49 = 1.02 \times 10^{-3}$ are marked in bold.

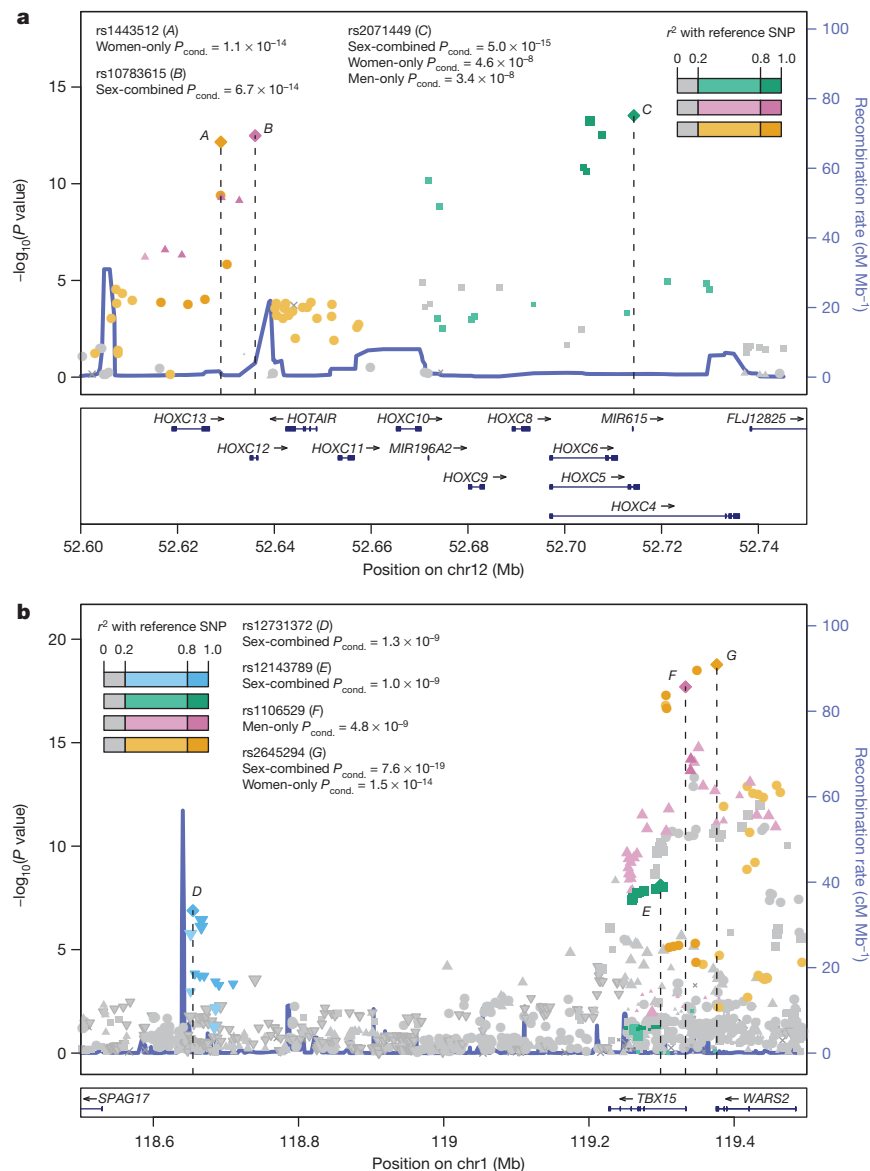
‡ Locus previously named *NISCH-STAB1*. Additional analyses that showed no significant evidence of heterogeneity between studies or due to ascertainment are provided in Supplementary Tables 27 and 28 (Supplementary Note).

(rs3936510) or only in men (rs459193; Extended Data Table 1, Supplementary Table 4). Other signals are more complex. The *TBX15-WARS2* locus showed different but correlated lead SNPs in men and women near *WARS2* ($r^2 = 0.43$), an independent signal near *TBX15*, and a distant independent signal near *SPAG17* (Fig. 1). At the *HOXC* gene cluster, conditional analyses identified independent ($r^2 < 0.01$) SNPs ~ 80 kb apart near *HOXC12-HOXC13-HOTAIR* and near *HOXC4-HOXC6* (Fig. 1). These results suggest that association signals mapping

to the same locus might act on different underlying genes and may not be relevant to the same sex.

We assessed the aggregate effects of the primary association signals at the 49 WHRadjBMI loci by calculating sex-combined and sex-specific risk based on genotypes of the lead SNPs. In a linear regression model, the risk scores were associated with WHRadjBMI, with a stronger effect in women than in men (overall effect per allele $\beta = 0.001$, $P = 6.7 \times 10^{-4}$, women $\beta = 0.002$, $P = 1.0 \times 10^{-11}$, men $\beta = 7.0 \times 10^{-4}$, $P = 0.02$;

Figure 1 | Regional SNP association plots illustrating the complex genetic architecture at two WHRadjBMI loci. **a, b,** Sex-combined meta-analysis SNP associations in European individuals were plotted with $-\log_{10} P$ values (left y axis) and estimated local recombination rate in blue (right y axis). Three index SNPs near *HOXC6-HOXC13* (denoted A–C) (**a**) and four near *TBX15-WARS2-SPAG17* (D–G) (**b**) were identified through approximate conditional analyses of sex-combined or sex-specific associations (values shown as $P_{\text{conditional}} < 5 \times 10^{-8}$, see Methods). The signals are distinguished by both colour and shape, and linkage disequilibrium (r^2) of nearby SNPs is shown by colour intensity gradient. Sample sizes for the index SNP associations are listed in Extended Data Table 1.



Extended Data Fig. 3 and Supplementary Note). The 49 SNPs explained 1.4% of the variance in WHRadjBMI overall, and more in women (2.4%) than in men (0.8%) (Supplementary Table 6). Compared to the 16 previously reported loci^{7,8}, the new loci almost doubled the explained variance in women and tripled that in men. We further estimated that the sex-combined variance explained by all HapMap SNPs¹⁵ (h^2_G) is 12.1% (s.e.m. = 2.9%).

At 17 loci with high-density coverage on the Metabochip¹¹, we used association summary statistics to define credible sets of SNPs with a high probability of containing a likely functional variant. The 99% credible sets at seven loci spanned < 20 kb, and at *HOXC13* included only a single noncoding SNP (Supplementary Table 7 and Supplementary Fig. 5). Imputation up to higher density reference panels will provide greater coverage and may have more potential to localize functional variants.

WHRadjBMI variants and other traits

Given the epidemiological correlations between central obesity and other anthropometric and cardiometabolic measures and diseases, we evaluated lead WHRadjBMI variants in association data from GWAS consortia for 22 traits. In total, 17 of the 49 variants were associated ($P < 5 \times 10^{-8}$) with at least one of the traits: high-density lipoprotein cholesterol (HDL; $n = 7$ SNPs), triglycerides ($n = 5$), low-density lipoprotein cholesterol (LDL; $n = 2$), adiponectin adjusted for BMI ($n = 3$),

fasting insulin adjusted for BMI ($n = 2$), T2D ($n = 1$), and height ($n = 7$) (Supplementary Tables 8 and 9). WHRadjBMI SNPs also showed enrichment for directional consistency among nominally significant ($P < 0.05$) associations with these traits and also with fasting and 2-h glucose, diastolic and systolic blood pressure, BMI and coronary artery disease (CAD) ($P_{\text{binomial}} < 0.05/23 = 0.0022$; Extended Data Table 2); these results were generally supported by meta-regression analysis of the regression coefficient estimates (Supplementary Table 10). Furthermore, our WHRadjBMI loci overlap with associations reported in the National Human Genome Research Institute (NHGRI) GWAS catalogue (Table 2 and Supplementary Table 11)¹⁶, the strongest of which is the locus near *LEKRI*, which is associated ($P = 2.0 \times 10^{-35}$) with birth weight¹⁷. Unsupervised hierarchical clustering of the corresponding matrix of association Z-scores showed three major clusters characterized by patterns of anthropometric and metabolic traits (Extended Data Fig. 4). These data extend knowledge about genetic links between WHRadjBMI and insulin-resistance-related traits; whether this reflects underlying causal relations between WHRadjBMI and these traits, or pleiotropic loci, cannot be inferred from our data.

Potential functional WHRadjBMI variants

We next examined variants in LD with the WHRadjBMI lead SNPs ($r^2 > 0.7$) for predicted effects on protein sequence, copy number, and

Table 2 | Candidate genes at new WHRadjBMI loci

SNP	Locus	eQTL ($P < 10^{-5}$)*	GRAIL ($P < 0.05$)†	DEPICT (FDR < 0.05)‡	Literature§	Other GWAS signals
rs905938	<i>DCST2</i>	ZBTB7B (PB, blood)	-	-	-	-
rs10919388	<i>GORAB</i>	-	-	-	-	-
rs1385167	<i>MEIS1</i>	-	-	-	<i>MEIS1</i>	-
rs1569135	<i>CALCRL</i>	-	<i>TFPI</i>	-	<i>CALCRL</i>	-
rs10804591	<i>PLXND1</i>	-	-	-	<i>PLXND1</i>	-
rs17451107	<i>LEKR1</i>	<i>TIPARP</i> (S,O), <i>LEKR1</i> (S)	-	-	-	Birth weight: <i>CCNL1</i> , <i>LEKR1</i>
rs3805389	<i>NMU</i>	-	-	-	<i>NMU</i>	-
rs9991328	<i>FAM13A</i>	<i>FAM13A</i> (S)	-	<i>FAM13A</i>	-	FI: <i>FAM13A</i>
rs303084	<i>SPATA5-FGF2</i>	-	<i>FGF2</i>	-	<i>FGF2</i> , <i>NUDT6</i> , <i>SPRY1</i>	-
rs9687846	<i>MAP3K1</i>	-	<i>MAP3K1</i>	-	<i>MAP3K1</i>	FI, TG: <i>ANKRD55</i> , <i>MAP3K1</i>
rs6556301	<i>FGFR4</i>	-	<i>MXD3</i>	-	<i>FGFR4</i>	Height
rs7759742	<i>BTNL2</i>	<i>HLA-DRA</i> (S), <i>KLHL31</i> (S)	-	(not analysed)	-	-
rs1776897	<i>HMGAI</i>	-	-	(not analysed)	<i>HMGAI</i>	Height: <i>HMGAI</i> , <i>C6orf106</i> , <i>LBH</i>
rs1534696	<i>SNX10</i>	<i>SNX10</i> (S), <i>CBX3</i> (S)	-	-	<i>SNX10</i>	-
rs7801581	<i>HOXA11</i>	-	<i>HOXA11</i>	<i>HOXA11</i>	<i>HOXA11</i>	-
rs7830933	<i>NKX2-6</i>	<i>STC1</i> (S)	-	-	<i>NKX2-6</i> , <i>STC1</i>	-
rs12679556	<i>MSC</i>	-	<i>EYA1</i>	<i>RP11-1102P16.1</i>	<i>MSC</i> , <i>EYA1</i>	-
rs10991437	<i>ABCA1</i>	-	-	-	<i>ABCA1</i>	-
rs7917772	<i>SFXN2</i>	-	-	-	<i>SFXN2</i>	Height
rs11231693	<i>MACROD1-VEGFB</i>	-	<i>VEGFB</i>	<i>MACROD1</i>	<i>MACROD1</i> , <i>VEGFB</i>	-
rs4765219	<i>CCDC92</i>	<i>CCDC92</i> (S, O, L), <i>ZNF664</i> (S, O)	<i>FAM101A</i>	-	-	Adiponectin, FI, HDL, TG: <i>CCDC92</i> , <i>ZNF664</i>
rs8042543	<i>KLF13</i>	-	<i>KLF13</i>	-	<i>KLF13</i>	-
rs8030605	<i>RFX7</i>	-	-	-	-	-
rs1440372	<i>SMAD6</i>	<i>SMAD6</i> (blood)	<i>SMAD6</i>	<i>SMAD6</i>	<i>SMAD6</i>	Height
rs2925979	<i>CMIP</i>	<i>CMIP</i> (S)	-	-	<i>CMIP</i> , <i>PLCG2</i>	Adiponectin, FI, HDL: <i>CMIP</i>
rs4646404	<i>PEMT</i>	-	-	<i>PEMT</i>	<i>PEMT</i>	-
rs8066985	<i>KCNJ2</i>	-	-	-	<i>KCNJ2</i>	-
rs12454712	<i>BCL2</i>	-	-	-	<i>BCL2</i>	-
rs12608504	<i>JUND</i>	<i>KIAA1683</i> (PB, O), <i>JUND</i> (LCL)	<i>JUND</i>	-	<i>JUND</i>	-
rs4081724	<i>CEBPA</i>	-	<i>CEBPA</i>	-	<i>CEBPA</i> , <i>CEBPG</i>	-
rs979012	<i>BMP2</i>	-	<i>BMP2</i>	<i>BMP2</i>	<i>BMP2</i>	Height: <i>BMP2</i>
rs224333	<i>GDF5</i>	<i>CEP250</i> (S, O), <i>UQCC</i> (blood, S, O, L, LCL)	<i>GDF5</i>	<i>GDF5</i>	<i>GDF5</i>	Height: <i>GDF5</i> , <i>UQCC</i>
rs6090583	<i>EYA2</i>	-	<i>EYA2</i>	<i>EYA2</i>	<i>EYA2</i>	-

Candidate genes based on secondary analyses or literature review. Details are provided in Supplementary Tables 8, 9, 11–13, 15, 19, 21 and Supplementary Note. The only non-synonymous variant in high LD with an index SNP was *GDF5* S276A. No copy number variants were identified. PB, peripheral blood mononuclear cells; FI, fasting insulin adjusted for BMI; HDL, high-density lipoprotein cholesterol; L, liver; LCL, lymphoblastoid cell line; O, omental adipose; S, subcutaneous adipose; TG, triglycerides.

* Gene transcript levels associated with the SNP in the indicated tissue(s).

† Genes in pathways identified as enriched by GRAIL analysis.

‡ Significant (FDR < 5%) pathway genes derived by DEPICT using GWAS-only results.

§ Most plausible candidate genes based on literature review.

|| Traits associated at $P < 5 \times 10^{-8}$ in GWAS or the GWAS catalogue using the index SNP or a proxy, and the genes(s) named.

cis-regulatory effects on expression (Table 2, Supplementary Tables 12–15 and Supplementary Note). At 11 of the new loci, lead WHRadjBMI SNPs were in LD with *cis*-expression quantitative trait loci (eQTLs) for transcripts in subcutaneous adipose tissue, omental adipose tissue, liver or blood cell types (Table 2 and Supplementary Table 15). No additional sex-specific eQTLs were identified, perhaps reflecting limited power (Supplementary Table 16).

At the 11 WHRadjBMI loci containing eQTLs, we compared the location of the candidate variants to regions of open chromatin (DNase I hypersensitivity and formaldehyde-assisted isolation of regulatory elements (FAIRE)) and histone modification enrichment (histone 3 Lys 4 methylation (H3K4me1), H3K4me2, H3K4me3, histone 3 Lys 27 acetylation (H3K27ac), and H3K9ac) in adipose, liver, skeletal muscle, bone, brain, blood and pancreatic islet tissues or cell lines (Supplementary Table 17). At 7 of these 11 loci, at least one variant was located in a putative regulatory element in two or more data sets from the same tissue as the eQTL, suggesting that these elements may influence transcriptional activity (Supplementary Table 18). For example, at *LEKR1*, five variants in LD with the WHRadjBMI lead SNP are located in a 1.1-kb region with evidence of enhancer activity (H3K4me1 and H3K27ac) in adipose tissue (Extended Data Fig. 5a).

We also examined whether any variants overlapped with open chromatin or histone modifications from only one of the tested tissues, possibly reflecting tissue-specific regulatory elements (Supplementary Table 18). For example, five variants in a 2.2-kb region, located 77 kb upstream from a *CALCRL* transcription start site, overlapped with peaks

in at least five data sets in endothelial cells (Extended Data Fig. 5b), suggesting that one or more of these variants may influence transcriptional activity. *CALCRL*, which is expressed in endothelial cells, is required for lipid absorption in the small intestine, and influences body weight in mice¹⁸. Other variants located in tissue-specific regulatory elements were detected at *NMU* for endothelial cells, at *KLF13* and *MEIS1* for liver, and at *GORAB* and *MSC* for bone (Supplementary Table 18).

Biological mechanisms

To identify potential functional connections between genes mapping to the 49 WHRadjBMI loci, we used three approaches (Supplementary Note). A survey of literature using GRAIL¹⁹ identified 15 genes with nominal significance ($P < 0.05$) for potential functional connectivity (Table 2 and Supplementary Table 19). The predefined gene set relationships across loci identified using MAGENTA²⁰ highlighted signalling pathways involving vascular endothelial growth factor (VEGF), phosphatase and tensin (PTEN) homologue, the insulin receptor, and peroxisome proliferator-activated receptors (Supplementary Table 20). VEGF signalling plays a central, complex role in angiogenesis, insulin resistance and obesity²¹, and PTEN signalling promotes insulin resistance²². Analyses using Data-driven Expression Prioritized Integration for Complex Traits (DEPICT)²³ facilitated prioritization of genes at associated loci, analyses of tissue specificity, and enrichment of reconstituted gene sets through integration of association results with expression data, protein–protein interactions, phenotypic data from gene knockout studies in mice, and predefined gene sets. DEPICT identified at least one

prioritized gene (false discovery rate (FDR) < 5%) at nine loci (Table 2 and Supplementary Table 21) and identified 234 reconstituted gene sets (161 after pruning of overlapping gene sets) enriched for genes at WHRadjBMI loci. Among these we highlight biologically plausible gene sets suggesting roles in body fat regulation (including adiponectin signalling, insulin sensitivity and regulation of glucose levels), skeletal growth, transcriptional regulation and development (Fig. 2 and Supplementary Table 22). We also note gene sets that are specific for abundance or development of metabolically active tissues including adipose, heart, liver and muscle. Specific genes at the loci were significantly enriched (FDR < 5%) for expression in adipocyte-related tissues, including abdominal subcutaneous fat (Fig. 2 and Supplementary Table 23). Together, these analyses identified processes related to insulin and adipose biology and highlight mesenchymal tissues, especially adipose tissue, as important to WHRadjBMI.

We also tested variants at the 49 WHRadjBMI loci for overlap with elements from 60 selected regulatory data sets from the ENCODE²⁴ and Epigenomic RoadMap²⁵ data and found evidence of enrichment in 12 data sets ($P < 0.05/60 = 8.3 \times 10^{-4}$; Extended Data Table 3). The strongest enrichments were detected for data sets typically attributed to enhancer activity (H3K4me1 and H3K27ac) in adipose, muscle, endothelial cells, and bone, suggesting that variants may regulate transcription in these tissues. These analyses point to mechanisms linking WHRadjBMI loci to regulation of gene expression in tissues highly relevant for adipocyte metabolism and insulin resistance.

We also reviewed functions of candidate genes located near new and previously established WHRadjBMI loci^{7,8}, identifying genes involved in adipogenesis, angiogenesis and transcriptional regulation (Table 2, literature review in the Supplementary Note). Adipogenesis candidate genes include *CEBPA*, *PPARG*, *BMP2*, *HOXC-mir196*, *SPRY1*, *TBX15*, and *PMT*. Of these, *CEBPA* and *PPARG* are essential for white adipose tissue differentiation²⁶, *BMP2* induces differentiation of mesenchymal

stem cells towards adipogenesis or osteogenesis²⁷, and *HOXC8* is a repressor of brown adipogenesis in mice that is regulated by miR-196a (ref. 28), also located within the *HOXC* region (Fig. 1). Angiogenesis genes may influence expansion and loss of adipose tissue²⁹; they include *VEGFA*, *VEGFB*, *RSPO3*, *STAB1*, *WARS2*, *PLXND1*, *MEIS1*, *FGF2*, *SMAD6* and *CALCL*. *VEGFB* is involved in endothelial targeting of lipids to peripheral tissues³⁰, and *PLXND1* limits blood vessel branching, antagonizes VEGF, and affects adipose inflammation^{31,32}. Transcriptional regulators at WHRadjBMI loci include *CEBPA*, *PPARG*, *MSC*, *SMAD6*, *HOXA*, *HOXC*, *ZBTB7B*, *JUND*, *KLF13*, *MEIS1*, *RFX7*, *NKX2-6* and *HMGAI*. Other candidate genes include *NMU*, *FGFR4* and *HMGAI*, for which mice deficient for the corresponding genes exhibit obesity, glucose intolerance and/or insulin resistance^{33–35}.

Five additional central obesity traits

To determine whether the WHRadjBMI variants exert their effects primarily through waist circumference or hip circumference and to identify loci that are not reported for WHRadjBMI, BMI or height^{36,37}, we performed association analyses for five additional traits: WCadjBMI, HIPadjBMI, WHR, waist circumference and hip circumference. On the basis of phenotypic data alone, waist circumference and hip circumference are highly correlated with BMI ($r = 0.59–0.92$), and WHR is highly correlated with WHRadjBMI ($r = 0.82–0.95$), while WCadjBMI and HIPadjBMI are moderately correlated with height ($r = 0.24–0.63$; Supplementary Table 24). In contrast to WHRadjBMI, which has almost no genetic correlation (see Methods) with height ($r_G < 0.04$; Extended Data Fig. 2c), WCadjBMI ($r_G = 0.42$) and HIPadjBMI ($r_G = 0.82$) have moderate genetic correlations with height. These data suggest that some, but not all, WCadjBMI and HIPadjBMI loci would be associated with height.

Across all meta-analyses, we identified an additional 19 loci associated with one of the five traits ($P < 5 \times 10^{-8}$), nine of which showed

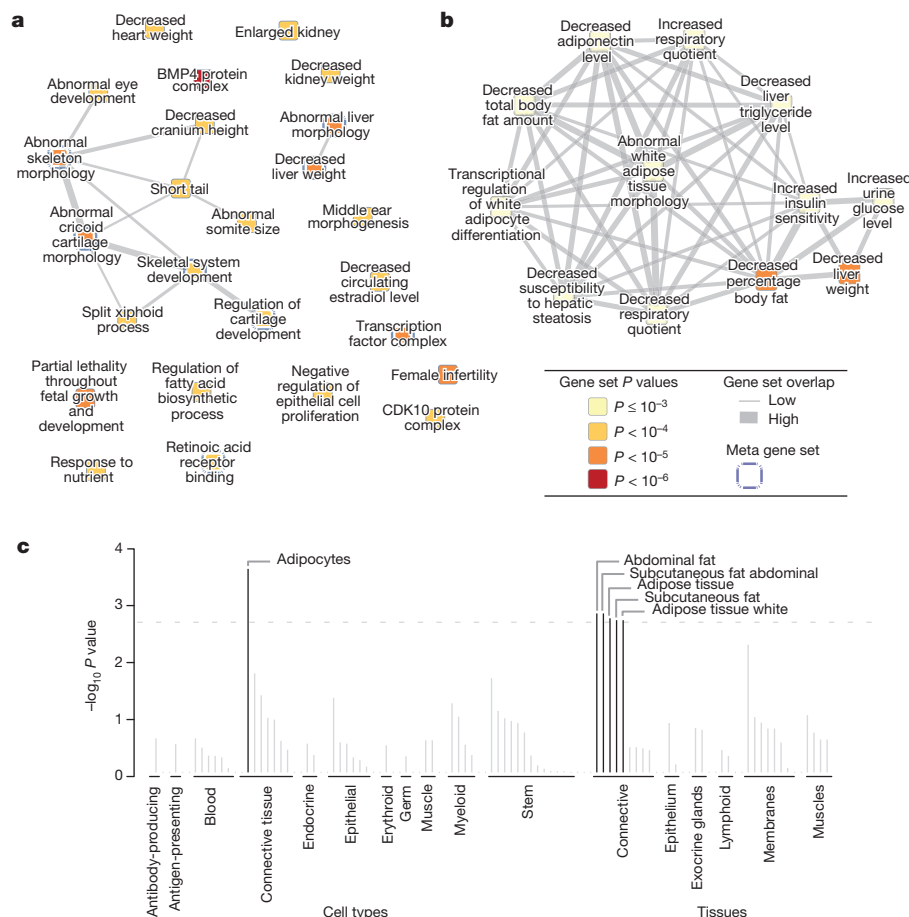


Figure 2 | Gene set enrichment and tissue expression of genes at WHRadjBMI-associated loci (GWAS-only $P < 10^{-5}$). **a**, Reconstituted gene sets found to be significantly enriched by DEPICT (FDR < 5%) are represented as nodes, with pairwise overlap denoted by the width of connecting lines and empirical enrichment P value indicated by colour intensity (darker is more significant). **b**, The ‘decreased liver weight’ meta-node, which consisted of 12 overlapping gene sets, including adiponectin signalling and insulin sensitivity. **c**, On the basis of expression patterns in 37,427 human microarray samples, annotations found to be significantly enriched by DEPICT are shown, grouped by type and significance.

Table 3 | New loci achieving genome-wide evidence of association ($P < 5 \times 10^{-8}$) with additional waist and hip circumference traits

SNP	Trait	Chr	Locus	EA*	EAF	Sex-combined			Women			Men			Sex diff.
						β	<i>P</i>	<i>N</i>	β	<i>P</i>	<i>N</i>	β	<i>P</i>	<i>N</i>	<i>P</i> †
Loci achieving genome-wide significance in European-ancestry meta-analyses															
rs10925060	WCadjBMI	1	OR2W5-NLRP3	T	0.03	0.017	2.2×10^{-5}	140,515	0.002	6.8×10^{-1}	85,186	0.045	9.1×10^{-13}	55,522	1.7×10^{-8}
rs10929925	HIP	2	SOX11	C	0.55	0.020	4.5×10^{-8}	207,648	0.021	9.0×10^{-6}	115,428	0.018	3.2×10^{-4}	92,499	6.1×10^{-1}
rs2124969	WCadjBMI	2	ITGB6	C	0.42	0.020	7.1×10^{-9}	231,284	0.016	3.5×10^{-4}	127,437	0.025	2.3×10^{-7}	104,039	1.4×10^{-1}
rs17472426	WCadjBMI	5	CCNJL	T	0.92	0.014	3.1×10^{-2}	217,564	-0.014	1.0×10^{-1}	119,804	0.052	4.3×10^{-8}	97,954	3.9×10^{-8}
rs7739232	HIPadjBMI	6	KLHL31	A	0.07	0.037	5.4×10^{-5}	131,877	0.063	1.0×10^{-8}	80,475	-0.004	7.5×10^{-1}	51,589	2.9×10^{-5}
rs13241538	HIPadjBMI	7	KLF14	C	0.48	0.017	1.6×10^{-6}	210,935	0.033	9.9×10^{-14}	117,210	-0.003	5.0×10^{-1}	93,911	2.0×10^{-9}
rs7044106	HIPadjBMI	9	C5	C	0.24	0.023	4.1×10^{-5}	143,412	0.039	5.7×10^{-9}	86,733	-0.003	6.9×10^{-1}	56,865	1.3×10^{-5}
rs11607976	HIP	11	MYEOV	C	0.70	0.022	4.2×10^{-8}	212,815	0.019	1.9×10^{-4}	118,391	0.024	7.7×10^{-6}	94,701	4.4×10^{-1}
rs1784203	WCadjBMI	11	KIAA1731	A	0.01	0.031	1.3×10^{-8}	63,892	0.000	9.9×10^{-1}	35,539	0.075	1.0×10^{-19}	28,353	1.2×10^{-1}
rs1394461	WHR	11	CNTN5	C	0.25	0.017	4.7×10^{-4}	144,349	0.035	3.6×10^{-8}	87,441	-0.011	1.6×10^{-1}	57,094	1.1×10^{-6}
rs319564	WHR	13	GPC6	C	0.45	0.014	3.4×10^{-5}	212,137	0.003	5.3E-01	117,970	0.027	1.6×10^{-8}	94,350	6.0×10^{-5}
rs2047937	WCadjBMI	16	ZNF423	C	0.50	0.019	4.7×10^{-8}	231,009	0.022	5.5×10^{-7}	127,288	0.014	3.6×10^{-3}	103,914	2.0×10^{-1}
rs2034088	HIPadjBMI	17	VPS53	T	0.53	0.021	4.8×10^{-9}	210,737	0.028	9.6×10^{-10}	117,142	0.014	6.5×10^{-3}	93,781	2.5×10^{-2}
rs1053593	HIPadjBMI	22	HMGXB4	T	0.65	0.021	3.9×10^{-8}	202,070	0.029	1.8×10^{-9}	114,347	0.011	5.1×10^{-2}	87,908	6.2×10^{-3}
Loci achieving genome-wide significance in all-ancestry meta-analyses															
rs1664789	WCadjBMI	5	ARL15	C	0.41	0.014	2.6×10^{-5}	244,110	0.005	2.8×10^{-1}	133,052	0.026	3.6×10^{-8}	109,025	4.4×10^{-4}
rs722585	HIPadjBMI	6	GMDS	G	0.68	0.015	2.1×10^{-4}	205,815	-0.001	8.8×10^{-1}	113,965	0.032	9.2×10^{-9}	89,831	4.3×10^{-6}
rs1144	WCadjBMI	7	SRPK2	C	0.34	0.019	3.1×10^{-8}	239,342	0.020	1.2×10^{-5}	131,398	0.018	4.1×10^{-4}	105,911	7.8×10^{-1}
rs2398893	WHR	9	PTPDC1	A	0.71	0.020	4.0×10^{-8}	226,572	0.019	5.1×10^{-5}	124,577	0.019	2.7×10^{-4}	99,968	9.5×10^{-1}
rs4985155‡	HIP	16	PDXDC1	A	0.66	0.018	4.5×10^{-7}	227,296	0.011	1.6×10^{-2}	125,048	0.029	9.7×10^{-9}	100,313	6.3×10^{-3}

P values and β coefficients for the association with the trait indicated in the meta-analysis of combined GWAS and MetaboChIP studies. The smallest P value for each SNP is shown in bold.

*The effect allele is the trait-increasing allele in the sex-combined analysis.

†Test for sex difference; values significant at the table-wide Bonferroni threshold of $0.05/19 = 2.63 \times 10^{-3}$ are marked in bold.

‡ $P = 7.3 \times 10^{-6}$ with height in ref. 43 (index SNP rs1136001; $r^2 = 0.79$, distance = 2,515 base pairs (bp)).

significantly larger effects ($P_{\text{difference}} < 0.05/19 = 0.003$) in one sex than in the other (Table 3, Supplementary Figs 1–4 and Supplementary Table 25). Three of four new loci with larger effects in women were associated with HIPadjBMI and three of five new loci with larger effects in men were associated with WCadjBMI. Most of the 19 loci showed some evidence of association with WHRadjBMI in sex-combined or sex-specific analyses, but four loci showed no association ($P > 0.01$) with WHRadjBMI, BMI, or height (Supplementary Tables 8 and 26).

We next asked whether the genes and pathways influencing these five traits are shared with WHRadjBMI or are distinct. Candidate genes were identified based on association with other traits, eQTLs, GRAIL and literature review (Extended Data Table 4 and Supplementary Tables 8, 11–13, 15–16 and 19). Candidate variants identified based on LD ($r^2 > 0.7$) included coding variants in *NTAN1* and *HMGXB4*, and six loci showed significant eQTLs in subcutaneous adipose tissue. On the basis of the literature, several candidate genes are involved in adipogenesis and insulin resistance. For example, delayed induction of preadipocyte transcription factor *ZNF423* in fibroblasts results in delayed adipogenesis³⁸, and *NLRP3* is part of inflammasome and pro-inflammatory T-cell populations in adipose tissue that contribute to inflammation and insulin resistance³⁹. GRAIL analyses identified connections that partially overlap with those identified for WHRadjBMI (Supplementary Table 19). Taken together, the additional loci appear to function in processes similar to the WHRadjBMI loci. The identification of loci that are more strongly associated with WCadjBMI or HIPadjBMI than the other anthropometric traits suggests that the additional traits characterize aspects of central obesity and fat distribution that are not captured by WHRadjBMI or BMI alone.

Discussion

These meta-analyses of GWAS and MetaboChIP data in up to 224,459 individuals identified additional loci associated with waist and hip circumference measures and help to determine the role of common genetic variation in body fat distribution that is distinct from BMI and height. Our results emphasize the strong sexual dimorphism in the genetic regulation of fat distribution traits, a characteristic not observed for overall obesity as assessed by BMI³⁶. Differences in body fat distribution between the sexes emerge in childhood, become more apparent during puberty⁴⁰, and change with menopause, generally attributed to the influence of sex

hormones^{41,42}. At loci with stronger effects in one sex than the other, these hormones may interact with transcription factors to regulate gene activity.

Annotation of the loci emphasized the role for mesenchymally derived tissues, especially adipose tissue, in fat distribution and central obesity. The development and regulation of adipose tissue deposition is closely associated with angiogenesis²⁹, a process highlighted by candidate genes at several WHRadjBMI loci. These tissues are implicated in insulin resistance, consistent with the enrichment of shared GWAS signals with lipids, T2D, and glycaemic traits. The identification of skeletal growth processes suggests that the underlying genes affect early development and/or differentiation of adipocytes from mesenchymal stem cells. By contrast, BMI has a substantial neuronal component, involving processes such as appetite regulation³⁶. Our results provide a foundation for future biological research in the regulation of body fat distribution and its connections with cardiometabolic traits, and offer potential target mechanisms for interventions in the risks associated with abdominal fat accumulation.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 20 November 2013; accepted 2 December 2014.

- Pischon, T. *et al.* General and abdominal adiposity and risk of death in Europe. *N. Engl. J. Med.* **359**, 2105–2120 (2008).
- Wang, Y., Rimm, E. B., Stampfer, M. J., Willett, W. C. & Hu, F. B. Comparison of abdominal adiposity and overall obesity in predicting risk of type 2 diabetes among men. *Am. J. Clin. Nutr.* **81**, 555–563 (2005).
- Canoy, D. Distribution of body fat and risk of coronary heart disease in men and women. *Curr. Opin. Cardiol.* **23**, 591–598 (2008).
- Snijder, M. B. *et al.* Associations of hip and thigh circumferences independent of waist circumference with the incidence of type 2 diabetes: the Hoorn Study. *Am. J. Clin. Nutr.* **77**, 1192–1197 (2003).
- Yusuf, S. *et al.* Obesity and the risk of myocardial infarction in 27,000 participants from 52 countries: a case-control study. *Lancet* **366**, 1640–1649 (2005).
- Mason, C., Craig, C. L. & Katzmarzyk, P. T. Influence of central and extremity circumferences on all-cause mortality in men and women. *Obesity* **16**, 2690–2695 (2008).
- Heid, I. M. *et al.* Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution. *Nature Genet.* **42**, 949–960 (2010).
- Randall, J. C. *et al.* Sex-stratified genome-wide association studies including 270,000 individuals show sexual dimorphism in genetic loci for anthropometric traits. *PLoS Genet.* **9**, e1003500 (2013).

9. Fox, C. S. *et al.* Genome-wide association of pericardial fat identifies a unique locus for ectopic fat. *PLoS Genet.* **8**, e1002705 (2012).
 10. Fox, C. S. *et al.* Genome-wide association for abdominal subcutaneous and visceral adipose reveals a novel locus for visceral fat in women. *PLoS Genet.* **8**, e1002695 (2012).
 11. Voight, B. F. *et al.* The metabochip, a custom genotyping array for genetic studies of metabolic, cardiovascular, and anthropometric traits. *PLoS Genet.* **8**, e1002793 (2012).
 12. Sanna, S. *et al.* Common variants in the GDF5-UQCC region are associated with variation in human height. *Nature Genet.* **40**, 198–203 (2008).
 13. Lango Allen, H. *et al.* Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* **467**, 832–838 (2010).
 14. Yang, J. *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nature Genet.* **44**, 369–375 (2012).
 15. Yang, J. *et al.* Common SNPs explain a large proportion of the heritability for human height. *Nature Genet.* **42**, 565–569 (2010).
 16. Hindorf, L. A. *et al.* A Catalog of Published Genome-Wide Association Studies. Available at <http://www.genome.gov/gwastudies>; accessed 31 January 2013.
 17. Freathy, R. M. *et al.* Variants in *ADCY5* and near *CCNL1* are associated with fetal growth and birth weight. *Nature Genet.* **42**, 430–435 (2010).
 18. Hoopes, S. L., Willcockson, H. H. & Caron, K. M. Characteristics of multi-organ lymphangiectasia resulting from temporal deletion of calcitonin receptor-like receptor in adult mice. *PLoS ONE* **7**, e45261 (2012).
 19. Raychaudhuri, S. *et al.* Identifying relationships among genomic disease regions: predicting genes at pathogenic SNP associations and rare deletions. *PLoS Genet.* **5**, e1000534 (2009).
 20. Segrè, A. V., Groop, L., Mootha, V. K., Daly, M. J. & Altshuler, D. Common inherited variation in mitochondrial genes is not enriched for associations with type 2 diabetes or related glycemic traits. *PLoS Genet.* **6**, e1001058 (2010).
 21. Elias, I., Franckhauser, S. & Bosch, F. New insights into adipose tissue VEGF-A actions in the control of obesity and insulin resistance. *Adipocyte* **2**, 109–112 (2013).
 22. Pal, A. *et al.* PTEN mutations as a cause of constitutive insulin sensitivity and obesity. *N. Engl. J. Med.* **367**, 1002–1011 (2012).
 23. Pers, T. *et al.* Biological interpretation of genome-wide association studies using predicted gene functions. *Nature Commun.* **6**, 5890 (2015).
 24. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
 25. Bernstein, B. E. *et al.* The NIH Roadmap Epigenomics Mapping Consortium. *Nature Biotechnol.* **28**, 1045–1048 (2010).
 26. Nakagami, H. The mechanism of white and brown adipocyte differentiation. *Diabetes Metab. J.* **37**, 85–90 (2013).
 27. Li, H. *et al.* miR-17–5p and miR-106a are involved in the balance between osteogenic and adipogenic differentiation of adipose-derived mesenchymal stem cells. *Stem Cell Res.* **10**, 313–324 (2013).
 28. Mori, M., Nakagami, H., Rodriguez-Araujo, G., Nimura, K. & Kaneda, Y. Essential role for miR-196a in brown adipogenesis of white fat progenitor cells. *PLoS Biol.* **10**, e1001314 (2012).
 29. Cao, Y. Angiogenesis and vascular functions in modulation of obesity, adipose metabolism, and insulin sensitivity. *Cell Metab.* **18**, 478–489 (2013).
 30. Hagberg, C. E. *et al.* Vascular endothelial growth factor B controls endothelial fatty acid uptake. *Nature* **464**, 917–921 (2010).
 31. Zygmunt, T. *et al.* Semaphorin-PlexinD1 signaling limits angiogenic potential via the VEGF decoy receptor sFlt1. *Dev. Cell* **21**, 301–314 (2011).
 32. Shimizu, I. *et al.* Semaphorin3E-induced inflammation contributes to insulin resistance in dietary obesity. *Cell Metab.* **18**, 491–504 (2013).
 33. Hanada, R. *et al.* Neuromedin U has a novel anorexigenic effect independent of the leptin signaling pathway. *Nature Med.* **10**, 1067–1073 (2004).
 34. Huang, X. *et al.* FGFR4 prevents hyperlipidemia and insulin resistance but underlies high-fat diet induced fatty liver. *Diabetes* **56**, 2501–2510 (2007).
 35. Foti, D. *et al.* Lack of the architectural factor HMGA1 causes insulin resistance and diabetes in humans and mice. *Nature Med.* **11**, 765–773 (2005).
 36. Locke, A. E. *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature* <http://dx.doi.org/10.1038/nature14177> (this issue).
 37. Wood, A. R. *et al.* Defining the role of common variation in the genomic and biological architecture of adult human height. *Nature Genet.* **46**, 1173–1186 (2014).
 38. Jäger, K. & Neuman, T. Human dermal fibroblasts exhibit delayed adipogenic differentiation compared with mesenchymal stem cells. *Stem Cells Dev.* **20**, 1327–1336 (2011).
 39. Goossens, G. H. *et al.* Expression of NLRP3 inflammasome and T cell population markers in adipose tissue are associated with insulin resistance and impaired glucose metabolism in humans. *Mol. Immunol.* **50**, 142–149 (2012).
 40. Maynard, L. M. *et al.* Childhood body composition in relation to body mass index. *Pediatrics* **107**, 344–350 (2001).
 41. Wells, J. C. Sexual dimorphism of body composition. *Best Pract. Res. Clin. Endocrinol. Metab.* **21**, 415–430 (2007).
 42. Lovejoy, J. C., Champagne, C. M., de Jonge, L., Xie, H. & Smith, S. R. Increased visceral fat and decreased energy expenditure during the menopausal transition. *Int. J. Obes.* **32**, 949–958 (2008).
 43. Okada, Y. *et al.* A genome-wide association study in 19,633 Japanese subjects identified LHX3-QSOX2 and IGF1 as adult height loci. *Hum. Mol. Genet.* **19**, 2303–2312 (2010).
- Supplementary Information** is available in the online version of the paper.
- Acknowledgements** We thank the more than 224,000 volunteers who participated in this study. Detailed acknowledgment of funding sources is provided in the Supplementary Note.
- Author Contributions** See the Supplementary Note for Author Contributions.
- Author Information** Summary results are available at <http://www.broadinstitute.org/collaboration/giant/>. Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to K.L.M. (mohlke@med.unc.edu) or C.M.L. (celi@well.ox.ac.uk).
- Dmitry Shungin^{1,2,3}, Thomas W. Winkler^{4*}, Damien C. Croteau-Chonka^{5,6*}, Teresa Ferreira^{7*}, Adam E. Locke^{8*}, Reedik Mägi^{9,10}, Rona J. Strawbridge¹⁰, Tine H. Pers^{1,12,13,14}, Krista Fischer⁹, Anne E. Justice¹⁵, Tsegaselassie Workalemahu¹⁶, Joseph M. W. Wu¹⁷, Martin L. Buchkovich¹⁸, Nancy L. Heard-Costa^{18,19}, Tamara S. Roman⁵, Alexander W. Drong²⁰, Ci Song^{20,21,22}, Stefan Gustafsson^{21,22}, Felix R. Day²³, Tonu Esko^{9,11,12,13}, Tove Fall^{20,21,22}, Zoltán Kutalik^{24,25,26}, Jian'an Luan²³, Joshua C. Randall^{7,27}, André Scherag^{28,29}, Sailaja Vedantam^{11,12}, Andrew R. Wood³⁰, Jin Chen³¹, Rudolf Fehrmann³², Juha Karjalainen³², Bratati Kahali³³, Ching-Ti Liu¹⁷, Ellen M. Schmidt³⁴, Devin Absher³⁵, Najaf Amin³⁶, Denise Anderson³⁷, Marian Beekman^{38,39}, Jennifer L. Bragg-Gresham⁴⁰, Steven Buyske^{41,42}, Ayse Demirkan^{36,43}, Georg B. Ehret^{44,45}, Mary F. Feitosa⁴⁶, Anuj Goel^{7,47}, Anne U. Jackson⁸, Toby Johnson^{25,26,48}, Marcus E. Kleber^{49,50}, Kati Kristiansson⁵¹, Massimo Mangino⁵², Irene Mateo Leach⁵³, Carolina Medina-Gomez^{54,55,56}, Cameron D. Palmer^{11,12}, Dorota Pasko³⁰, Sonali Pechlivanis²⁸, Marjolijn J. Peters^{54,56}, Inga Prokopenko^{57,58}, Alena Stančáková⁵⁹, Yun Ju Sung⁶⁰, Toshiaki Tanaka⁶¹, Alexander Teumer⁶², Jana V. Van Vliet-Ostapchouk⁶³, Loïc Yengo^{64,65,66}, Weihua Zhang^{67,68}, Eva Albrecht⁶⁹, Johan Ärnlöv^{21,22,70}, Gillian M. Arscott⁷¹, Stefania Bandinelli⁷², Amy Barrett⁷³, Claire Bellis^{73,74}, Amanda J. Bennett⁵⁷, Christian Berne⁷⁵, Matthias Blüher^{76,77}, Stefan Böhringer^{38,78}, Fabrice Bonnet⁷⁹, Yvonne Böttcher⁷⁶, Marcel Bruinenberg⁸⁰, Delia B. Carba⁸¹, Ida H. Caspersen⁸², Robert Clarke⁸³, E. Warwick Daw⁴⁶, Joris Deelen^{38,39}, Ewa Deelman⁸⁴, Graciela Delgado⁴⁹, Alex S. F. Doney⁸⁵, Niina Eklund^{51,86}, Michael R. Erdos⁸⁷, Karol Estrada^{12,56,88}, Elodie Eury^{64,65,66}, Nele Friedrich⁸⁹, Melissa E. Garcia⁹⁰, Vilmantas Giedraitis⁹¹, Bruna Gigante⁹², Alan S. Go⁹³, Alain Gola⁹⁴, Harald Grallert^{69,95,96}, Tanja B. Grammer⁴⁹, Jürgen Gräßler⁹⁷, Jagvir Grewal^{67,68}, Christopher J. Groves⁵⁷, Toomas Haller⁹, Goran Hallmans⁹⁸, Catharina A. Hartman⁹⁹, Maija Hassinen¹⁰⁰, Caroline Hayward¹⁰¹, Kauko Heikkilä¹⁰², Karl-Heinz Herzog^{103,104,105}, Quinta Helmer^{38,78,106}, Hans L. Hillege^{53,107}, Oddgeir Holmen¹⁰⁸, Steven C. Hunt¹⁰⁹, Aaron Isaacs^{36,110}, Till Ittermann¹¹¹, Alan L. James^{112,113}, Ingegerd Johansson³, Thorhildur Juliusdottir⁹¹, Ioanna Panagiotou Kalafati¹¹⁴, Leena Kinnunen⁵¹, Wolfgang Koenig⁵⁰, Ishminder K. Kooner⁶⁷, Wolfgang Kratzer¹¹⁵, Claudia Lamina¹¹⁶, Karin Leander⁹², Nanette R. Lee⁸¹, Peter Lichtner¹¹⁷, Lars Lind¹¹⁸, Jaana Lindström⁵¹, Stéphane Lobbens^{64,65,66}, Mattias Lorentzen¹¹⁹, François Mach⁴⁵, Patrik K. E. Magnusson²⁰, Anubha Mahajan⁷, Wendy L. McArdle¹²⁰, Cristina Menni⁵², Sigrun Merger¹²¹, Evelin Mihailov⁹¹, Lili Milani⁹², Rebecca Mills⁶⁷, Alireza Moayyeri^{52,123}, Keri L. Monda^{15,124}, Simon P. Mooijaart^{38,125}, Thomas W. Mühleisen^{126,127}, Antonella Mulas¹²⁸, Gabriele Müller¹²⁹, Martina Müller-Nurasyid^{69,130,131,132}, Ramaiah Nagaraja¹³³, Michael A. Nalls¹³⁴, Narisu Narisu⁸⁷, Nicola Glorioso¹³⁵, Ilja M. Nolte¹⁰⁷, Matthias Olden⁴, Nigel W. Rayner^{7,27,57}, Frida Renstrom², Janina S. Ried⁶⁹, Neil R. Robertson^{7,57}, Lynda M. Rose¹³⁶, Serena Sanna¹²⁸, Hubert Scharnagl¹³⁷, Salome Scholtens⁸⁰, Bengt Sennblad^{10,138}, Thomas Seufferlein¹¹⁵, Colleen M. Sifton¹³⁹, Albert Vernon Smith^{140,141}, Kathleen Stirrups^{27,142}, Heather M. Stringham⁸, Johan Sundström¹¹⁸, Morris A. Swertz³², Amy J. Swift⁸⁷, Ann-Christine Syvänen^{21,143}, Bamidele O. Tayo¹⁴⁴, Barbara Thorand^{96,145}, Gudmar Thorleifsson¹⁴⁶, Andreas Tomaschitz¹⁴⁷, Chiara Troffa¹³⁵, Floor V. A. van Oort¹⁴⁸, Niek Verweij⁵³, Judith M. Vonk¹⁰⁷, Lindsay L. Waite³⁵, Roman Wennauer¹⁴⁹, Tom Wilsaard¹⁵⁰, Mary K. Wojczynski¹⁴⁶, Andrew Wong¹⁵¹, Qunyan Zhang⁴⁶, Jing Hua Zhao²³, Eoin P. Brennan¹⁵², Murim Choi¹⁵³, Per Eriksson¹⁰, Lasse Folkersen¹⁰, Anders Franco-Cereceda¹⁵⁴, Ali G. Gharavi¹⁵⁵, Åsa K. Hedman^{7,21,22}, Marie-France Hivert^{156,157}, Jinyan Huang^{158,159}, Stavroula Kanoni¹⁴², Fredrik Karpe^{57,160}, Sarah Keildson⁷, Krzysztof Kiryluk¹⁵⁵, Liming Liang^{159,161}, Richard P. Lifton¹⁶², Baoshan Ma^{159,163}, Amy J. McKnight¹⁶⁴, Ruth McPherson¹⁶⁵, Andres Metspalu^{91,122}, Josine L. Min¹²⁰, Miriam F. Moffatt¹⁶⁶, Grant W. Montgomery¹⁶⁷, Joanne M. Murabito^{18,168}, George Nicholson^{169,170}, Dale R. Nyholt^{167,171}, Christian Olsson¹⁵⁴, John R. B. Perry^{7,30,52}, Eva Reinmaa⁹, Rany M. Salem^{11,12,13}, Niina Sandholm^{172,173,174}, Eric E. Schadt¹⁷⁵, Robert A. Scott²³, Lisette Stolk^{38,56}, Edgar E. Vallejo¹⁷⁶, Harm-Jan Westra³², Krina T. Zondervan^{7,177}, The ADIPOGen Consortium†, The CARDIOGRAMplusC4D Consortium†, The CKDGen Consortium†, The GEOS Consortium†, The GENIE Consortium†, The GLGC†, The ICBP†, The International Endogene Consortium†, The LifeLines Cohort Study†, The MAGIC Investigator†, The MuTHER Consortium†, The PAGE Consortium†, The ReproGen Consortium†, Philippe Amouyel¹⁷⁸, Dominique Arveiler¹⁷⁹, Stephan J. L. Bakker¹⁸⁰, John Beilby^{71,181}, Richard N. Bergman¹⁸², John Blangero⁷³, Morris J. Brown¹⁸³, Michel Burnier¹⁸⁴, Harry Campbell¹⁸⁵, Aravinda Chakravarti⁴⁴, Peter S. Chines⁸⁷, Simone Claudi-Boehm¹²¹, Francis S. Collins⁸⁷, Dana C. Crawford^{186,187}, John Danesh¹⁸⁸, Ulf de Faire⁹², Eco J. C. de Geus^{189,190}, Marcus Dörr^{191,192}, Raimund Erbel¹⁹³, Johan G. Eriksson^{51,194,195}, Martin Farrall^{7,47}, Ele Ferrannini^{196,197}, Jean Ferrières¹⁹⁸, Nita G. Forouhi²³, Terrence Forrester¹⁹⁹, Oscar H. Franco^{54,55}, Ron T. Gansevoort¹⁸⁰, Christian Gieger⁶⁹, Vilhelmur Gudnason^{140,141}, Christopher A. Haiman²⁰⁰, Tamara B. Harris⁹⁰, Andrew T. Hattersley²⁰¹, Markku Heliövaara⁵¹, Andrew A. Hicks²⁰², Aaron D. Hingorani²⁰³, Wolfgang Hoffmann^{111,192}, Albert Hofman^{54,55}, Georg Homuth⁶², Steve E. Humphries²⁰⁴, Elina

Hyppönen^{205,206,207,208}, Thomas Illig^{95,209}, Marjo-Riitta Jarvelin^{68,105,210,211,212,213}, Berit Johansen⁸², Pekka Jousilahti⁵¹, Antti M. Jula⁵¹, Jaakko Kaprio^{51,86,102}, Frank Kee²¹⁴, Sirkka M. Keinanen-Kiukkaanniemi^{215,216}, Jaspal S. Kooner^{67,166,217}, Charles Kooperberg²¹⁸, Peter Kovacs^{76,77}, Aldi T. Kraja⁴⁶, Meena Kumari^{219,220}, Kari Kuusamäa⁵¹, Johanna Kuusisto²²¹, Timo A. Lakka^{100,222,223}, Claudia Langenberg^{23,219}, Loic Le Marchand²²⁴, Terho Lehtimäki²²⁵, Valeriya Lysenko^{226,227}, Satu Männistö⁵¹, André Marette^{228,229}, Tara C. Matise⁴², Colin A. McKenzie¹⁹⁹, Barbara McKnight²³⁰, Arthur W. Musk²³¹, Stefan Möhlenkamp¹⁹³, Andrew D. Morris⁸⁵, Mari Nelis⁹, Claes Ohlsson¹¹⁹, Albertine J. Oldehinkel⁹⁹, Ken K. Ong^{23,151}, Lyle J. Palmer^{232,233}, Brenda W. Penninx^{190,234}, Annette Peters^{95,132,145}, Peter P. Pramstaller^{202,235}, Olli T. Raitakar^{236,237}, Tuomo Rankinen²³⁸, D. C. Rao^{46,60,239}, Treva K. Rice^{60,239}, Paul M. Ridker^{136,240}, Marylyn D. Ritchie²⁴¹, Igor Rudan^{185,242}, Veikko Salomaa⁵¹, Nilesh J. Samani²⁴³, Jouko Saramies²⁴⁵, Mark A. Sarzynski^{97,246}, Peter E. H. Schwarz^{97,246}, Alan R. Shuldiner^{247,248,249}, Jan A. Staessen^{250,251}, Valgerdur Steinthorsdottir¹⁴⁶, Ronald P. Stolk¹⁰⁷, Konstantin Strauch^{69,131}, Anke Tönjes^{76,77}, Angelo Tremblay²⁵², Elena Tremoli²⁵³, Marie-Claude Vohl^{225,254}, Uwe Völker^{62,192}, Peter Vollenweider²⁵⁵, James F. Wilson¹⁸⁵, Jacqueline C. Witterman⁵⁵, Linda S. Adair²⁵⁶, Murielle Bochud^{257,258}, Bernhard O. Boehm^{259,260}, Stefan R. Bornstein⁹⁷, Claude Bouchard²³⁸, Stéphane Cauchi^{64,65,66}, Mark J. Caulfield²⁶¹, John C. Chambers^{67,68,217}, Daniel I. Chasman^{136,240}, Richard S. Cooper¹⁴⁴, George Dedoussis¹¹⁴, Luigi Ferrucci⁶¹, Philippe Froguel^{58,64,65,66}, Hans-Jörgen Grabe^{262,263}, Anders Hamsten¹⁰, Jennie Hui^{71,181,264}, Kristian Hveem¹⁰⁸, Karl-Heinz Jöckel²⁸, Mika Kivimäki²¹⁹, Diana Kuh¹⁵¹, Markku Laakso²²¹, Yongmei Liu²⁶⁵, Winfried März^{49,137,266}, Patricia B. Munroe²⁶¹, Inger Njølstad¹⁵⁰, Ben A. Oostra^{36,110,267}, Colin N. A. Palmer⁸⁵, Nancy L. Pedersen²⁰, Markus Perola^{9,51,86}, Louis Pérusse^{229,252}, Ulrike Peters²¹⁸, Chris Power²⁰⁸, Thomas Quertermous²⁶⁸, Rainer Rauramaa^{100,223}, Fernando Rivadeneira^{54,55,56}, Timo E. Saaristo^{269,270}, Danish Saleheen^{188,271,272}, Juha Sinisalo²⁷³, P. Elina Slagboom^{38,39}, Harold Snieder¹⁰⁷, Tim D. Spector⁵², Unnur Thorsteinsdottir^{146,274}, Michael Stumvoll^{76,77}, Jaakko Tuomilehto^{51,275,276,277}, André G. Uitterlinden^{54,55,56}, Matti Uusitupa^{78,279}, Pim van der Harst^{52,53,280}, Giovanni Veronesi²⁸¹, Mark Walker²⁸², Nicholas J. Wareham²³, Hugh Watkins^{7,47}, H-Erich Wichmann^{283,284,285}, Goncalo R. Abecasis⁸, Themistocles L. Assimes²⁶⁸, Sonja I. Berndt²⁸⁶, Michael Boehnke⁸, Ingrid B. Borecki⁴⁶, Panos Deloukas^{27,142,287}, Lude Franke³², Timothy M. Frayling³⁰, Leif C. Groop^{86,227}, David J. Hunter^{6,16,159}, Robert C. Kaplan²⁸⁸, Jeffrey R. O'Connell^{146,274}, Lu Qi^{6,16}, David Schlessinger¹³³, David P. Strachan²⁸⁹, Kari Stefansson^{146,274}, Cornelia M. van Duijn^{36,54,55,110}, Cristen J. Willer^{31,34,290}, Peter M. Visscher^{291,292}, Jian Yang^{291,292}, Joel N. Hirschhorn^{11,12,13}, M. Carola Zillikens^{54,56}, Mark I. McCarthy³, Elizabeth K. Speliotes³³, Kari E. North^{15,294}, Caroline S. Fox¹⁸, Inês Barroso^{27,295,296}, Paul W. Franks^{12,16}, Erik Ingelsson^{7,21,22}, Iris M. Heid^{4,69}, Ruth J. F. Loos^{23,297,298,299}, S. L. Adrienne Cupples^{17,18}, Andrew P. Morris^{7,9,300}, Cecilia M. Lindgren^{7,12} & Karen L. Mohlke⁵

¹Department of Public Health and Clinical Medicine, Unit of Medicine, Umeå University, 901 87 Umeå, Sweden. ²Department of Clinical Sciences, Genetic & Molecular Epidemiology Unit, Lund University Diabetes Center, Skåne University Hospital, 205 02 Malmö, Sweden. ³Department of Odontology, Umeå University, 901 85 Umeå, Sweden. ⁴Department of Genetic Epidemiology, Institute of Epidemiology and Preventive Medicine, University of Regensburg, D-93053 Regensburg, Germany. ⁵Department of Genetics, University of North Carolina, Chapel Hill, North Carolina 27599, USA. ⁶Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts 02115, USA. ⁷Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK. ⁸Center for Statistical Genetics, Department of Biostatistics, University of Michigan, Ann Arbor, Michigan 48109, USA. ⁹Estonian Genome Center, University of Tartu, Tartu 51010, Estonia. ¹⁰Atherosclerosis Research Unit, Center for Molecular Medicine, Department of Medicine, Karolinska Institutet, Stockholm 17176, Sweden. ¹¹Divisions of Endocrinology and Genetics and Center for Basic and Translational Obesity Research, Boston Children's Hospital, Boston, Massachusetts 02115, USA. ¹²Broad Institute of the Massachusetts Institute of Technology and Harvard University, Cambridge, Massachusetts 02142, USA. ¹³Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA. ¹⁴Center for Biological Sequence Analysis, Department of Systems Biology, Technical University of Denmark, Lyngby 2800, Denmark. ¹⁵Department of Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA. ¹⁶Department of Nutrition, Harvard School of Public Health, Boston, Massachusetts 02115, USA. ¹⁷Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts 02118, USA. ¹⁸National Heart, Lung, and Blood Institute, the Framingham Heart Study, Framingham Massachusetts 01702, USA. ¹⁹Department of Neurology, Boston University School of Medicine, Boston, Massachusetts 02118, USA. ²⁰Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm 17177, Sweden. ²¹Science for Life Laboratory, Uppsala University, Uppsala 75185, Sweden. ²²Department of Medical Sciences, Molecular Epidemiology, Uppsala University, Uppsala 75185, Sweden. ²³MRC Epidemiology Unit, University of Cambridge School of Clinical Medicine, Institute of Metabolic Science, Cambridge Biomedical Campus, Cambridge CB2 0QQ, UK. ²⁴Institute of Social and Preventive Medicine (IUMSP), Centre Hospitalier Universitaire Vaudois (CHUV), Lausanne 1010, Switzerland. ²⁵Swiss Institute of Bioinformatics, Lausanne 1015, Switzerland. ²⁶Department of Medical Genetics, University of Lausanne, Lausanne 1005, Switzerland. ²⁷Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1SA, UK. ²⁸Institute for Medical Informatics, Biometry and Epidemiology (IMBE), University Hospital Essen, Essen, 45147 Germany. ²⁹Clinical Epidemiology, Integrated Research and Treatment Center, Center for Sepsis Control and Care (CSCC), Jena University Hospital, Jena 07743, Germany. ³⁰Genetics of Complex Traits, University of Exeter Medical School, University of Exeter, Exeter EX1 2LU, UK. ³¹Department of Internal Medicine, Division of Cardiovascular Medicine, University of Michigan, Ann Arbor, Michigan 48109, USA. ³²Department of Genetics, University Medical Center Groningen, University of Groningen, 9700 RB Groningen, The Netherlands. ³³Department of Internal Medicine, Division of Gastroenterology, and

Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, Michigan 48109, USA. ³⁴Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, Michigan 48109, USA. ³⁵HudsonAlpha Institute for Biotechnology, Huntsville, Alabama 35806, USA. ³⁶Genetic Epidemiology Unit, Department of Epidemiology, Erasmus MC University Medical Center, 3015 GE Rotterdam, The Netherlands. ³⁷Telethon Institute for Child Health Research, Centre for Child Health Research, The University of Western Australia, Perth, Western Australia 6008, Australia. ³⁸Netherlands Consortium for Healthy Aging (NCHA), Leiden University Medical Center, Leiden 2300 RC, The Netherlands. ³⁹Department of Molecular Epidemiology, Leiden University Medical Center, 2300 RC Leiden, The Netherlands. ⁴⁰Kidney Epidemiology and Cost Center, University of Michigan, Ann Arbor, Michigan 48109, USA. ⁴¹Department of Statistics & Biostatistics, Rutgers University, Piscataway, New Jersey 08854, USA. ⁴²Department of Genetics, Rutgers University, Piscataway, New Jersey 08854, USA. ⁴³Department of Human Genetics, Leiden University Medical Center, 2333 ZC Leiden, The Netherlands. ⁴⁴Center for Complex Disease Genomics, McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA. ⁴⁵Cardiology, Department of Specialties of Internal Medicine, Geneva University Hospital, Geneva 1211, Switzerland. ⁴⁶Department of Genetics, Washington University School of Medicine, St Louis, Missouri 63110, USA. ⁴⁷Division of Cardiovascular Medicine, Radcliffe Department of Medicine, University of Oxford, Oxford OX3 9DU, UK. ⁴⁸University Institute for Social and Preventative Medicine, Centre Hospitalier Universitaire Vaudois (CHUV), University of Lausanne, Lausanne 1005, Switzerland. ⁴⁹Vth Department of Medicine (Nephrology, Hypertensiology, Endocrinology, Diabetology, Rheumatology), Medical Faculty of Mannheim, University of Heidelberg, D-68187 Mannheim, Germany. ⁵⁰Department of Internal Medicine II, Ulm University Medical Center, D-89081 Ulm, Germany. ⁵¹National Institute for Health and Welfare, FI-00271 Helsinki, Finland. ⁵²Department of Twin Research and Genetic Epidemiology, King's College London, London SE1 7EH, UK. ⁵³Department of Cardiology, University Medical Center Groningen, University of Groningen, 9700RB Groningen, The Netherlands. ⁵⁴Netherlands Consortium for Healthy Aging (NCHA), 3015GE Rotterdam, The Netherlands. ⁵⁵Department of Epidemiology, Erasmus MC University Medical Center, 3015GE Rotterdam, The Netherlands. ⁵⁶Department of Internal Medicine, Erasmus MC University Medical Center, 3015GE Rotterdam, The Netherlands. ⁵⁷Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Oxford OX3 7LJ, UK. ⁵⁸Department of Genomics of Common Disease, School of Public Health, Imperial College London, Hammersmith Hospital, London W12 0NN, UK. ⁵⁹University of Eastern Finland, FI-70210 Kuopio, Finland. ⁶⁰Division of Biostatistics, Washington University School of Medicine, St Louis, Missouri 63110, USA. ⁶¹Translational Gerontology Branch, National Institute on Aging, Baltimore, Maryland 21225, USA. ⁶²Interfaculty Institute for Genetics and Functional Genomics, University Medicine Greifswald, D-17475 Greifswald, Germany. ⁶³Department of Endocrinology, University of Groningen, University Medical Center Groningen, Groningen, 9700 RB, The Netherlands. ⁶⁴CNRS UMR 8199, F-59019 Lille, France. ⁶⁵European Genomic Institute for Diabetes, F-59000 Lille, France. ⁶⁶Université de Lille 2, F-59000 Lille, France. ⁶⁷Ealing Hospital NHS Trust, Middlesex UB1 3HW, UK. ⁶⁸Department of Epidemiology and Biostatistics, Imperial College London, London W2 1PG, UK. ⁶⁹Institute of Genetic Epidemiology, Helmholtz Zentrum München - German Research Center for Environmental Health, D-85764 Neuherberg, Germany. ⁷⁰School of Health and Social Studies, Dalarna University, SE-791 88 Falun, Sweden. ⁷¹PathWest Laboratory Medicine of Western Australia, Nedlands, Western Australia 6009, Australia. ⁷²Geriatric Unit, Azienda Sanitaria Firenze (ASF), 50125 Florence, Italy. ⁷³Department of Genetics, Texas Biomedical Research Institute, San Antonio, Texas 78227, USA. ⁷⁴Genomics Research Centre, Institute of Health and Biomedical Innovation, Queensland University of Technology, Brisbane, Queensland 4001, Australia. ⁷⁵Department of Medical Sciences, Endocrinology, Diabetes and Metabolism, Uppsala University, Uppsala 75185, Sweden. ⁷⁶Integrated Research and Treatment Center (IFB) Adiposity Diseases, University of Leipzig, D-04103 Leipzig, Germany. ⁷⁷Department of Medicine, University of Leipzig, D-04103 Leipzig, Germany. ⁷⁸Department of Medical Statistics and Bioinformatics, Leiden University Medical Center, 2300 RC Leiden, The Netherlands. ⁷⁹Inserm UMR991, Department of Endocrinology, University of Rennes, F-35000 Rennes, France. ⁸⁰LifeLines Cohort Study, University Medical Center Groningen, University of Groningen, 9700 RB Groningen, The Netherlands. ⁸¹USC-Office of Population Studies Foundation, Inc., University of San Carlos, Cebu City 6000, Philippines. ⁸²Department of Biology, Norwegian University of Science and Technology, 7491 Trondheim, Norway. ⁸³Clinical Trial Service Unit and Epidemiological Studies Unit, Nuffield Department of Population Health, University of Oxford, Oxford OX3 7LF, UK. ⁸⁴Information Sciences Institute, University of Southern California, Marina del Rey, California 90292, USA. ⁸⁵Medical Research Institute, University of Dundee, Ninewells Hospital and Medical School, Dundee DD1 9SY, UK. ⁸⁶Institute for Molecular Medicine, University of Helsinki, FI-00014 Helsinki, Finland. ⁸⁷Medical Genomics and Metabolic Genetics Branch, National Human Genome Research Institute, NIH, Bethesda, Maryland 20892, USA. ⁸⁸Analytic and Translational Genetics Unit, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts 02114, USA. ⁸⁹Institute of Clinical Chemistry and Laboratory Medicine, University Medicine Greifswald, D-17475 Greifswald, Germany. ⁹⁰Laboratory of Epidemiology and Population Sciences, National Institute on Aging, NIH, Bethesda, Maryland 20892, USA. ⁹¹Department of Public Health and Caring Sciences, Geriatrics, Uppsala University, Uppsala 75185, Sweden. ⁹²Division of Cardiovascular Epidemiology, Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden, Stockholm 17177, Sweden. ⁹³Kaiser Permanente, Division of Research, Oakland, California 94612, USA. ⁹⁴Service of Therapeutic Education for Diabetes, Obesity and Chronic Diseases, Geneva University Hospital, Geneva CH-1211, Switzerland. ⁹⁵Research Unit of Molecular Epidemiology, Helmholtz Zentrum München - German Research Center for Environmental Health, D-85764 Neuherberg, Germany. ⁹⁶German Center for Diabetes Research (DZD), D-85764 Neuherberg, Germany. ⁹⁷Department of Medicine III, University Hospital Carl Gustav Carus, Technische Universität Dresden, D-01307 Dresden, Germany. ⁹⁸Department of Public Health and Clinical Medicine, Unit of Nutritional Research, Umeå University, Umeå 90187, Sweden. ⁹⁹Department of Psychiatry, University of Groningen, University Medical Center

Groningen, 9700RB Groningen, The Netherlands. ¹⁰⁰Kuopio Research Institute of Exercise Medicine, FI-70100 Kuopio, Finland. ¹⁰¹MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Western General Hospital, Edinburgh EH4 2XU, UK. ¹⁰²Hjelt Institute Department of Public Health, University of Helsinki, FI-00014 Helsinki, Finland. ¹⁰³Institute of Biomedicine, University of Oulu, FI-90014 Oulu, Finland. ¹⁰⁴Medical Research Center Oulu and Oulu University Hospital, FI-90014 Oulu, Finland. ¹⁰⁵Biocenter Oulu, University of Oulu, FI-90014 Oulu, Finland. ¹⁰⁶Faculty of Psychology and Education, VU University Amsterdam, 1081BT Amsterdam, The Netherlands. ¹⁰⁷Department of Epidemiology, University Medical Center Groningen, University of Groningen, 9700 RB Groningen, The Netherlands. ¹⁰⁸Department of Public Health and General Practice, Norwegian University of Science and Technology, Trondheim 7489, Norway. ¹⁰⁹Cardiovascular Genetics Division, Department of Internal Medicine, University of Utah, Salt Lake City, Utah 84108, USA. ¹¹⁰Center for Medical Systems Biology, 2300 RC Leiden, The Netherlands. ¹¹¹Institute for Community Medicine, University Medicine Greifswald, D-17475 Greifswald, Germany. ¹¹²Department of Pulmonary Physiology and Sleep Medicine, Nedlands, Western Australia 6009, Australia. ¹¹³School of Medicine and Pharmacology, University of Western Australia, Crawley 6009, Australia. ¹¹⁴Department of Dietetics-Nutrition, Harokopio University, 17671 Athens, Greece. ¹¹⁵Department of Internal Medicine I, Ulm University Medical Centre, D-89081 Ulm, Germany. ¹¹⁶Division of Genetic Epidemiology, Department of Medical Genetics, Molecular and Clinical Pharmacology, Innsbruck Medical University, 6020 Innsbruck, Austria. ¹¹⁷Institute of Human Genetics, Helmholtz Zentrum München - German Research Center for Environmental Health, D-85764 Neuherberg, Germany. ¹¹⁸Department of Medical Sciences, Cardiovascular Epidemiology, Uppsala University, Uppsala 75185, Sweden. ¹¹⁹Centre for Bone and Arthritis Research, Department of Internal Medicine and Clinical Nutrition, Institute of Medicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg 413 45, Sweden. ¹²⁰School of Social and Community Medicine, University of Bristol, Bristol BS8 2BN, UK. ¹²¹Division of Endocrinology, Diabetes and Metabolism, Ulm University Medical Centre, D-89081 Ulm, Germany. ¹²²Institute of Molecular and Cell Biology, University of Tartu, Tartu 51010, Estonia. ¹²³Farr Institute of Health Informatics Research, University College London, London NW1 2DA, UK. ¹²⁴The Center for Observational Research, Amgen, Inc., Thousand Oaks, California 91320, USA. ¹²⁵Department of Gerontology and Geriatrics, Leiden University Medical Center, 2300 RC Leiden, The Netherlands. ¹²⁶Department of Genomics, Life & Brain Center, University of Bonn, 53127 Bonn, Germany. ¹²⁷Institute of Human Genetics, University of Bonn, 53127 Bonn, Germany. ¹²⁸Istituto di Ricerca Genetica e Biomedica (IRGB), Consiglio Nazionale delle Ricerche, Cagliari, Sardinia 09042, Italy. ¹²⁹Center for Evidence-based Healthcare, University Hospital Carl Gustav Carus, Technische Universität Dresden, D-01307 Dresden, Germany. ¹³⁰Department of Medicine I, University Hospital Grosshadern, Ludwig-Maximilians-Universität, D-81377 Munich, Germany. ¹³¹Institute of Medical Informatics, Biometry and Epidemiology, Chair of Genetic Epidemiology, Ludwig-Maximilians-Universität, D-81377 Munich, Germany. ¹³²Deutsches Forschungszentrum für Herz-Kreislauferkrankungen (DZHK) (German Research Center for Cardiovascular Research), Munich Heart Alliance, D-80636 Munich, Germany. ¹³³Laboratory of Genetics, National Institute on Aging, Baltimore, Maryland 21224, USA. ¹³⁴Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, Maryland 20892, USA. ¹³⁵Hypertension and Related Diseases Centre - AOU, University of Sassari Medical School, Sassari 07100, Italy. ¹³⁶Division of Preventive Medicine, Brigham and Women's Hospital, Boston, Massachusetts 02215, USA. ¹³⁷Clinical Institute of Medical and Chemical Laboratory Diagnostics, Medical University of Graz, Graz 8036, Austria. ¹³⁸Science for Life Laboratory, Karolinska Institutet, Stockholm 171 65, Sweden. ¹³⁹Department of Medicine, University of Washington, Seattle, Washington 98101, USA. ¹⁴⁰Icelandic Heart Association, Kopavogur 201, Iceland. ¹⁴¹University of Iceland, Reykjavik 101, Iceland. ¹⁴²William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London EC1M 6BQ, UK. ¹⁴³Department of Medical Sciences, Molecular Medicine, Uppsala University, Uppsala 75144, Sweden. ¹⁴⁴Department of Public Health Sciences, Stritch School of Medicine, Loyola University of Chicago, Maywood, Illinois 61053, USA. ¹⁴⁵Institute of Epidemiology II, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany, D-85764 Neuherberg, Germany. ¹⁴⁶deCODE Genetics, Amgen Inc., Reykjavik 101, Iceland. ¹⁴⁷Department of Cardiology, Medical University of Graz, Graz 8036, Austria. ¹⁴⁸Department of Child and Adolescent Psychiatry, Psychology, Erasmus MC University Medical Centre, 3000 CB Rotterdam, The Netherlands. ¹⁴⁹Department of Clinical Chemistry, Ulm University Medical Centre, D-89081 Ulm, Germany. ¹⁵⁰Department of Community Medicine, Faculty of Health Sciences, UiT The Arctic University of Norway, 9037 Tromsø, Norway. ¹⁵¹MRC Unit for Lifelong Health and Ageing at University College London, London WC1B 5JU, UK. ¹⁵²Diabetes Complications Research Centre, Conway Institute, School of Medicine and Medical Sciences, University College Dublin, Dublin 4, Ireland. ¹⁵³Department of Biomedical Sciences, Seoul National University College of Medicine, Seoul 110-799, Korea. ¹⁵⁴Cardiothoracic Surgery Unit, Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm 17176, Sweden. ¹⁵⁵Department of Medicine, Columbia University College of Physicians and Surgeons, New York 10032, USA. ¹⁵⁶Department of Population Medicine, Harvard Pilgrim Health Care Institute, Harvard Medical School, Boston, Massachusetts 02215, USA. ¹⁵⁷Massachusetts General Hospital, Boston, Massachusetts 02114, USA. ¹⁵⁸State Key Laboratory of Medical Genomics, Shanghai Institute of Hematology, Rui Jin Hospital Affiliated with Shanghai Jiao Tong University School of Medicine, Shanghai 200025, China. ¹⁵⁹Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts 02115, USA. ¹⁶⁰NHR Oxford Biomedical Research Centre, OUH Trust, Oxford OX3 7LE, UK. ¹⁶¹Harvard School of Public Health, Department of Biostatistics, Harvard University, Boston, Massachusetts 02115, USA. ¹⁶²Department of Genetics, Howard Hughes Medical Institute, Yale University School of Medicine, New Haven, New Haven, Connecticut 06520, USA. ¹⁶³College of Information Science and Technology, Dalian Maritime University, Dalian, Liaoning 116026, China. ¹⁶⁴Nephrology Research, Centre for Public Health, Queen's University of Belfast, Belfast, County Down BT9 7AB, UK. ¹⁶⁵University of Ottawa Heart Institute, Ottawa K1Y 4W7, Canada. ¹⁶⁶National Heart and Lung Institute, Imperial College London, London SW3 6LY, UK. ¹⁶⁷QIMR Berghofer Medical Research Institute, Brisbane, Queensland 4006, Australia. ¹⁶⁸Section of General Internal Medicine, Boston University School of Medicine, Boston, Massachusetts 02118, USA. ¹⁶⁹Department of Statistics, University of Oxford, 1 South Parks Road, Oxford OX1 3TG, UK. ¹⁷⁰MRC Harwell, Harwell Science and Innovation Campus, Harwell OX11 0QG, UK. ¹⁷¹Institute of Health and Biomedical Innovation, Queensland University of Technology, Brisbane, Queensland 4059, Australia. ¹⁷²Department of Biomedical Engineering and Computational Science, Aalto University School of Science, FI-00076 Helsinki, Finland. ¹⁷³Department of Medicine, Division of Nephrology, Helsinki University Central Hospital, FI-00290 Helsinki, Finland. ¹⁷⁴Folkhälsan Institute of Genetics, Folkhälsan Research Center, FI-00290 Helsinki, Finland. ¹⁷⁵Icahn Institute for Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, New York 10580, USA. ¹⁷⁶Computer Science Department, Tecnológico de Monterrey, Atizapán de Zaragoza, 52926, Mexico. ¹⁷⁷Nuffield Department of Obstetrics & Gynaecology, University of Oxford, Oxford OX3 7BN, UK. ¹⁷⁸Institut Pasteur de Lille; INSERM, U744; Université de Lille 2; F-59000 Lille, France. ¹⁷⁹Department of Epidemiology and Public Health, EA4330, University of Strasbourg, Faculty of Medicine, Strasbourg, France. ¹⁸⁰Department of Internal Medicine, University Medical Center Groningen, University of Groningen, 9700RB Groningen, The Netherlands. ¹⁸¹Pathology and Laboratory Medicine, The University of Western Australia, Perth, Western Australia 6009, Australia. ¹⁸²Cedars-Sinai Diabetes and Obesity Research Institute, Los Angeles, California 90048, USA. ¹⁸³Clinical Pharmacology Unit, University of Cambridge, Addenbrooke's Hospital, Hills Road, Cambridge CB2 2QQ, UK. ¹⁸⁴Service of Nephrology, Department of Medicine, Lausanne University Hospital (CHUV), Lausanne 1005, Switzerland. ¹⁸⁵Centre for Population Health Sciences, University of Edinburgh, Teviot Place, Edinburgh EH8 9AG, UK. ¹⁸⁶Center for Human Genetics Research, Vanderbilt University Medical Center, Nashville, Tennessee 37203, USA. ¹⁸⁷Department of Molecular Physiology and Biophysics, Vanderbilt University, Nashville, Tennessee 37232, USA. ¹⁸⁸Department of Public Health and Primary Care, University of Cambridge, Cambridge CB1 8RN, UK. ¹⁸⁹Biological Psychology, VU University Amsterdam, 1081BT Amsterdam, The Netherlands. ¹⁹⁰Institute for Research in Extramural Medicine, Institute for Health and Care Research, VU University, 1081BT Amsterdam, The Netherlands. ¹⁹¹Department of Internal Medicine B, University Medicine Greifswald, D-17475 Greifswald, Germany. ¹⁹²DZHK (Deutsches Zentrum für Herz-Kreislaufforschung - German Centre for Cardiovascular Research), partner site Greifswald, D-17475 Greifswald, Germany. ¹⁹³Clinic of Cardiology, West-German Heart Centre, University Hospital Essen, 45122 Essen, Germany. ¹⁹⁴Department of General Practice and Primary Health Care, University of Helsinki, FI-00290 Helsinki, Finland. ¹⁹⁵Unit of General Practice, Helsinki University Central Hospital, Helsinki FI-00290, Finland. ¹⁹⁶Department of Internal Medicine, University of Pisa, Pisa 56100, Italy. ¹⁹⁷National Research Council Institute of Clinical Physiology, University of Pisa, Pisa 56124, Italy. ¹⁹⁸Department of Cardiology, Toulouse University School of Medicine, Rangueil Hospital, 31400 Toulouse, France. ¹⁹⁹UWI Solutions for Developing Countries, The University of the West Indies, Mona, Kingston 7, Jamaica. ²⁰⁰Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, California 90089, USA. ²⁰¹Institute of Biomedical & Clinical Science, University of Exeter, Barrack Road, Exeter EX2 5DW, UK. ²⁰²Center for Biomedicine, European Academy Bozen, Bolzano (EURAC), Bolzano 39100, Italy (affiliated Institute of the University of Lübeck, D-23562 Lübeck, Germany). ²⁰³Institute of Cardiovascular Science, University College London, London WC1E 6BT, UK. ²⁰⁴Centre for Cardiovascular Genetics, Institute Cardiovascular Sciences, University College London, London WC1E 6JJ, UK. ²⁰⁵Sansom Institute for Health Research, University of South Australia, Adelaide 5000, South Australia, Australia. ²⁰⁶School of Population Health, University of South Australia, Adelaide 5000, South Australia, Australia. ²⁰⁷South Australian Health and Medical Research Institute, Adelaide 5000, South Australia, Australia. ²⁰⁸Population, Policy, and Practice, University College London Institute of Child Health, London WC1N 1EH, UK. ²⁰⁹Hannover Unified Biobank, Hannover Medical School, Hannover, D-30625 Hannover, Germany. ²¹⁰National Institute for Health and Welfare, FI-90101 Oulu, Finland. ²¹¹MRC Health Protection Agency (HPA) Centre for Environment and Health, School of Public Health, Imperial College London, London W2 1PG, UK. ²¹²Unit of Primary Care, Oulu University Hospital, FI-90220 Oulu, Finland. ²¹³Institute of Health Sciences, University of Oulu, FI-90014 Oulu, Finland. ²¹⁴UK Clinical Research Collaboration Centre of Excellence for Public Health (NI), Queens University of Belfast, Belfast BT7 1NN, Northern Ireland, UK. ²¹⁵Institute of Health Sciences, Faculty of Medicine, University of Oulu, FI-90014 Oulu, Finland. ²¹⁶Unit of Primary Health Care/General Practice, Oulu University Hospital, FI-90220 Oulu, Finland. ²¹⁷Imperial College Healthcare NHS Trust, London W12 0HS, UK. ²¹⁸Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington 98109, USA. ²¹⁹Department of Epidemiology and Public Health, University College London, London WC1E 6BT, UK. ²²⁰Department of Biological and Social Epidemiology, University of Essex, Wivenhoe Park, Colchester, Essex CO4 3SQ, UK. ²²¹Department of Medicine, Kuopio University Hospital and University of Eastern Finland, FI-70210 Kuopio, Finland. ²²²Department of Physiology, Institute of Biomedicine, University of Eastern Finland, Kuopio Campus, FI-70211 Kuopio, Finland. ²²³Department of Clinical Physiology and Nuclear Medicine, Kuopio University Hospital and University of Eastern Finland, FI-70210 Kuopio, Finland. ²²⁴Epidemiology Program, University of Hawaii Cancer Center, Honolulu, Hawaii 96813, USA. ²²⁵Department of Clinical Chemistry, Fimlab Laboratories and School of Medicine University of Tampere, FI-33520 Tampere, Finland. ²²⁶Steno Diabetes Center A/S, Gentofte DK-2820, Denmark. ²²⁷Lund University Diabetes Centre and Department of Clinical Science, Diabetes & Endocrinology Unit, Lund University, Malmö 221 00, Sweden. ²²⁸Institut Universitaire de Cardiologie et de Pneumologie de Québec, Faculty of Medicine, Laval University, Quebec QC G1V 0A6, Canada. ²²⁹Institute of Nutrition and Functional Foods, Laval University, Quebec QC G1V 0A6, Canada. ²³⁰Department of Biostatistics, University of Washington, Seattle, Washington 98195, USA. ²³¹Department of Respiratory Medicine, Sir Charles Gairdner Hospital, Nedlands, Western Australia 6009, Australia. ²³²Epidemiology and Obstetrics & Gynaecology, University of Toronto, Toronto, Ontario M5G 1E2, Canada. ²³³Genetic Epidemiology & Biostatistics Platform, Ontario Institute for Cancer Research, Toronto, Ontario M5G 0A3, Canada. ²³⁴Department

of Psychiatry, Neuroscience Campus, VU University Amsterdam, 1081 BT Amsterdam, The Netherlands. ²³⁵Department of Neurology, General Central Hospital, Bolzano 39100, Italy. ²³⁶Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, FI-20521 Turku, Finland. ²³⁷Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, FI-20521 Turku, Finland. ²³⁸Human Genomics Laboratory, Pennington Biomedical Research Center, Baton Rouge, Louisiana 70808, USA. ²³⁹Department of Psychiatry, Washington University School of Medicine, St Louis, Missouri 63110, USA. ²⁴⁰Harvard Medical School, Boston, Massachusetts 02115, USA. ²⁴¹Center for Systems Genomics, The Pennsylvania State University, University Park, Pennsylvania 16802, USA. ²⁴²Croatian Centre for Global Health, Faculty of Medicine, University of Split, 21000 Split, Croatia. ²⁴³Department of Cardiovascular Sciences, University of Leicester, Glenfield Hospital, Leicester LE3 9QP, UK. ²⁴⁴National Institute for Health Research (NIHR) Leicester Cardiovascular Biomedical Research Unit, Glenfield Hospital, Leicester LE3 9QP, UK. ²⁴⁵South Carelia Central Hospital, 53130 Lappeenranta, Finland. ²⁴⁶Paul Langerhans Institute Dresden, German Center for Diabetes Research (DZD), 01307 Dresden, Germany. ²⁴⁷Division of Endocrinology, Diabetes and Nutrition, University of Maryland School of Medicine, Baltimore, Maryland 21201, USA. ²⁴⁸Program for Personalized and Genomic Medicine, University of Maryland School of Medicine, Baltimore, Maryland 21201, USA. ²⁴⁹Geriatric Research and Education Clinical Center, Veterans Administration Medical Center, Baltimore, Maryland 21201, USA. ²⁵⁰Department of Epidemiology, Maastricht University, 6229 HA Maastricht, The Netherlands. ²⁵¹Research Unit Hypertension and Cardiovascular Epidemiology, KU Leuven Department of Cardiovascular Sciences, University of Leuven, B-3000 Leuven, Belgium. ²⁵²Department of Kinesiology, Laval University, Quebec, QC G1V 0A6, Canada. ²⁵³Dipartimento di Scienze Farmacologiche e Biomolecolari, Università di Milano & Centro Cardiologico Monzino, Istituto di Ricovero e Cura a Carattere Scientifico, Milan 20133, Italy. ²⁵⁴Department of Food Science and Nutrition, Laval University, Quebec, QC G1V 0A6, Canada. ²⁵⁵Department of Internal Medicine, University Hospital (CHUV) and University of Lausanne, 1011, Switzerland. ²⁵⁶Department of Nutrition, University of North Carolina, Chapel Hill, North Carolina 27599, USA. ²⁵⁷Institute of Social and Preventive Medicine (IUMSP), Centre Hospitalier Universitaire Vaudois and University of Lausanne, 1010 Lausanne, Switzerland. ²⁵⁸Ministry of Health, Victoria, Republic of Seychelles. ²⁵⁹Lee Kong Chian School of Medicine, Imperial College London and Nanyang Technological University, Singapore, 637553 Singapore, Singapore. ²⁶⁰Department of Internal Medicine I, Ulm University Medical Centre, D-89081 Ulm, Germany. ²⁶¹Department of Clinical Pharmacology, William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London EC1M 6BQ, UK. ²⁶²Department of Psychiatry and Psychotherapy, University Medicine Greifswald, HELIOS-Hospital Stralsund, D-17475 Greifswald, Germany. ²⁶³German Center for Neurodegenerative Diseases (DZNE), Rostock, Greifswald, D-17475 Greifswald, Germany. ²⁶⁴School of Population Health, The University of Western Australia, Nedlands, Western Australia 6009, Australia. ²⁶⁵Center for Human Genetics, Division of Public Health Sciences, Wake Forest School of Medicine, Winston-Salem, North Carolina 27157, USA. ²⁶⁶Synlab Academy, Synlab Services GmbH, 68163 Mannheim, Germany. ²⁶⁷Department of Clinical Genetics, Erasmus MC University Medical Center, 3000 CA Rotterdam, The Netherlands. ²⁶⁸Department of Medicine, Stanford University School of Medicine, Palo Alto, California 94304, USA. ²⁶⁹Finnish Diabetes Association, Kirjoniementie 15, FI-33680 Tampere, Finland. ²⁷⁰Pirkanmaa Hospital District, FI-33521 Tampere, Finland. ²⁷¹Center for Non-Communicable Diseases, Karachi, Pakistan. ²⁷²Department of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104 USA. ²⁷³Helsinki University Central Hospital Heart and Lung Center, Department of Medicine, Helsinki University Central Hospital, FI-00290 Helsinki, Finland. ²⁷⁴Faculty of Medicine, University of Iceland, Reykjavik 101, Iceland. ²⁷⁵Instituto de Investigación Sanitaria del Hospital Universitario La Paz (IdiPAZ), 28046 Madrid, Spain. ²⁷⁶Diabetes Research Group, King Abdulaziz University, 21589 Jeddah, Saudi Arabia. ²⁷⁷Centre for Vascular Prevention, Danube-University Krems, 3500 Krems, Austria. ²⁷⁸Department of Public Health and Clinical Nutrition, University of Eastern Finland, FI-70211 Kuopio, Finland. ²⁷⁹Research Unit, Kuopio University Hospital, FI-70210 Kuopio, Finland. ²⁸⁰Durrer Center for Cardiogenetic Research, Interuniversity Cardiology Institute Netherlands-Netherlands Heart Institute, 3501 DG Utrecht, The Netherlands. ²⁸¹EPIMED Research Center, Department of Clinical and Experimental Medicine, University of Insubria, Varese I-21100, Italy. ²⁸²Institute of Cellular Medicine, Newcastle University, Newcastle NE1 7RU, UK. ²⁸³Institute of Medical Informatics, Biometry and Epidemiology, Chair of Epidemiology, Ludwig-Maximilians-Universität, D-85764 Munich, Germany. ²⁸⁴Klinikum Grosshadern, D-81377 Munich, Germany. ²⁸⁵Institute of Epidemiology I, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany, D-85764 Neuherberg, Germany. ²⁸⁶Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, USA. ²⁸⁷Princess Al-Jawhara Al-Brahim Centre of Excellence in Research of Hereditary Disorders (PACER-HD), King Abdulaziz University, 21589 Jeddah, Saudi Arabia. ²⁸⁸Albert Einstein College of Medicine, Department of Epidemiology and Population Health, Belfer 1306, New York 10461, USA. ²⁸⁹Division of Population Health Sciences & Education, St George's, University of London, London SW17 0RE, UK. ²⁹⁰Department of Human Genetics, University of Michigan, Ann Arbor, Michigan 48109, USA. ²⁹¹Queensland Brain Institute, The University of Queensland, Brisbane 4072, Australia. ²⁹²The University of Queensland Diamantina Institute, The Translation Research Institute, Brisbane 4012, Australia. ²⁹³Oxford NIHR Biomedical Research Centre, Oxford University Hospitals NHS Trust, Oxford OX3 7LJ, UK. ²⁹⁴Carolina Center for Genome Sciences, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA. ²⁹⁵University of Cambridge Metabolic Research Laboratories, Institute of Metabolic Science, Addenbrooke's Hospital, Cambridge CB2 0QQ, UK. ²⁹⁶NIHR Cambridge Biomedical Research Centre, Institute of Metabolic Science, Addenbrooke's Hospital, Cambridge CB2 0QQ, UK. ²⁹⁷The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA. ²⁹⁸The Genetics of Obesity and Related Metabolic Traits Program, The Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA. ²⁹⁹The Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA. ³⁰⁰Department of Biostatistics, University of Liverpool, Liverpool L69 3GA, UK.

†A list of authors and affiliations appears in the Supplementary Information.

*These authors contributed equally to this work.

§These authors jointly supervised this work.

METHODS

Study overview. Our study included 224,459 individuals of European, east Asian, south Asian and African-American ancestry. The European ancestry arm included 142,762 individuals from 57 cohorts genotyped with genome-wide SNP arrays and 67,326 individuals from 44 cohorts genotyped with the MetaboChip¹¹ (Extended Data Fig. 1 and Supplementary Table 1). The non-European ancestry arm comprised ~1,700 individuals from one cohort of east Asian ancestry, ~3,400 individuals from one cohort of south Asian ancestry, and ~9,200 individuals from six cohorts of African-American ancestry, all genotyped with the MetaboChip. There was no overlap between individuals genotyped with genome-wide SNP arrays and MetaboChip. For each study, local institutional committees approved study protocols and confirmed that informed consent was obtained. No statistical methods were used to predetermine sample size.

Traits. Our primary trait was WHRadjBMI, the ratio of waist and hip circumferences adjusted for age, age-squared, study-specific covariates if necessary and BMI. For each cohort, residuals were calculated for men and women separately and then transformed by the inverse standard normal function. Cohorts with related men and women provided inverse standard normal transformed sex-combined residuals. For each cohort, the same transformations were applied to other traits: (1) WHR without adjustment for BMI (WHR); (2) waist circumference with (WCadjBMI) and without adjustment for BMI; and (3) hip circumference with (HIPadjBMI) and without adjustment for BMI.

European ancestry meta-analysis for genome-wide SNP array data. Sample and SNP quality control were undertaken within each cohort⁴⁴ (Supplementary Table 3). The GWAS scaffold in each cohort was imputed up to CEU haplotypes from HapMap resulting in ~2.5 million SNPs. Each directly typed and imputed SNP passing quality control was tested for association with each trait under an additive model in a linear regression framework (Supplementary Table 3). SNP positions are reported based on NCBI Build 36. For each cohort, sex-specific association summary statistics were corrected for residual population structure using the genomic control inflation factor⁴⁵ (median $\lambda_{GC} = 1.01$, range = 0.99–1.08). SNPs were removed before meta-analysis if they had a minor allele count ≤ 3 , deviation from Hardy–Weinberg equilibrium exact $P < 10^{-6}$, directly genotyped SNP call rate $< 95\%$, or low imputation quality (below 0.3 for MACH, 0.4 for IMPUTE, and 0.8 for PLINK). Association summary statistics for each trait were combined via inverse-variance weighted fixed-effects meta-analysis and corrected for a second round of genomic control to account for structure between cohorts (Extended Data Fig. 1 and Supplementary Fig. 1).

European ancestry meta-analysis for MetaboChip data. Sample and SNP quality control analyses were undertaken in each cohort (Supplementary Table 3). Each SNP passing quality control was tested for association with each trait under an additive model using linear regression. The MetaboChip array¹¹ is enriched, by design, for loci associated with anthropometric and cardiometabolic traits, thus, we based our correction on 4,425 SNPs selected for inclusion based on associations with QT-interval that were not expected to be associated with anthropometric traits (> 500 kb from variants on MetaboChip⁴⁶ for these traits). These study-specific inflation factors had a median $\lambda_{GC} = 1.01$ (range 0.93–1.11), with only one study exceeding 1.10. After removing SNPs for quality control as described in the previous section, association summary statistics were combined via inverse-variance weighted fixed-effects meta-analysis and corrected for a second round of genomic control on the basis of QT-interval SNPs to account for structure between cohorts.

European ancestry meta-analyses. Association summary statistics from the two parts of the European ancestry arm were combined via inverse-variance weighted fixed-effects meta-analysis using METAL⁴⁷ with no further genomic control correction. Results were reported for SNPs with a sex-combined sample size $\geq 50,000$. The meta-analyses were repeated for men and women separately for each trait. Analyses were corrected for population structure within each sex. The meta-analysis of WHRadjBMI in men included up to 93,480 individuals, and in women up to 116,742 individuals.

Meta-analyses of studies of all ancestries. Sample and SNP quality control, tests of association, genomic control correction (median $\lambda_{GC} = 1.01$, range = 0.90–1.17, with only one study exceeding 1.10), and meta-analyses were performed as described above. Association summary statistics from the European and non-European ancestry meta-analyses were combined via inverse-variance weighted fixed-effects meta-analysis without further genomic control correction.

Heterogeneity. For each lead SNP, we tested for sex differences based on the sex-specific beta estimates and standard errors, while accounting for potential correlation between estimates as previously described¹⁰. Similarly, we tested for potential differences in effects between European and non-European samples, comparing the effects from GWAS + MetaboChip data for Europeans and MetaboChip data for non-Europeans, and we tested for differences between population-based studies and samples ascertained on diabetes status, and cardiovascular disease, or both. In assessing effects of ascertainment overall, we compared effects in seven subsets of our

study sample using population-based studies (that is, those not ascertained on any phenotype) as the reference population: (1) all studies ascertained on any phenotype, (2) T2D cases, (3) T2D controls, (4) T2D cases plus controls, (5) CAD cases, (6) CAD controls and (7) CAD cases plus controls. We evaluated significance for heterogeneity tests within each comparison using a Bonferroni-corrected P value of $0.05/49 = 0.05/49 = 1.02 \times 10^{-3}$ as well as an FDR threshold⁴⁸ of $< 5\%$ (Supplementary Table 28). Between-study heterogeneity in all meta-analyses was assessed using I^2 statistics⁴⁹.

Heritability and genetic and phenotypic correlations of waist traits. We calculated the heritability and genetic correlations of several central obesity traits using variance component models^{50,51} in the Framingham Heart Study (FHS) and TwinGene study. In this approach, the phenotypic variance is decomposed into additive genetic, non-additive genetic, and environmental sources of variation (including model error), and for sets of traits, the covariances between traits. We report narrow sense heritability (h^2), the ratio of the additive genetic variance to the total phenotypic variance. Sex-specific inverse normal trait residuals, adjusted for age (and cohort in FHS), were used to estimate heritability separately in men and women, using variance components analysis in SOLARv.4.2.7 (FHS)⁵² or Mx1.703 (TwinGene)⁵³. Additionally, the sex-specific residuals were used to conduct bivariate quantitative variance component genetic analyses that calculate genetic and environmental correlations between traits. The genetic correlations obtained are estimates of the additive effects of shared genes, and a genetic correlation significantly different from zero suggests a direct influence of the same genes on more than one trait. Similarly, significant environmental correlations suggest shared environmental effects.

We estimated sex-stratified correlations between all waist traits, as well as BMI, height, and weight in TwinGene, FHS, KORA and EGCUT. In TwinGene and FHS, age-adjusted Pearson correlations were used; in EGCUT and KORA, correlations were adjusted for age and age-squared.

European ancestry approximate conditional analyses. To evaluate the evidence for multiple association signals within identified loci, we performed approximate conditional analyses of sex-combined, women-specific and men-specific data as implemented in the GCTA software^{14,54}. This approach makes use of association summary statistics from the combined European ancestry meta-analysis and a reference data set of individual-level genotype data to estimate LD between variants and hence also the approximate correlation between allelic effect estimates in a joint association model.

To evaluate robustness of the GCTA results, we performed analyses using two reference data sets: Prospective Investigation of the Vasculature in Uppsala Seniors (PIVUS) consisting of 949 individuals from Uppsala County, Sweden, with both GWAS and MetaboChip genotype data; and Atherosclerosis Risk in Communities (ARIC) consisting of 6,654 individuals of European descent from four communities in the United States with GWAS data. Both GWAS data sets were imputed using data from phase II of the International HapMap Project⁵⁵. Results shown use the PIVUS reference data set because MetaboChip genotypes are available (see a comparison in the Supplementary Note). Assuming that the LD correlations between SNPs more than 10 megabases (Mb) away are zero, and using each reference data set in turn, we performed a genome-wide stepwise selection procedure to select associated SNPs one-by-one at a $P < 5 \times 10^{-8}$. For each locus at which multiple association signals were observed in the sex-combined, women-, and/or men-specific data, the SNPs selected by GCTA as independently associated with WHRadjBMI in any of the three meta-analyses are reported, with the SNP identified in the sex-combined analysis taken by default when proxies are identified in the women- and/or men-specific analyses. For SNPs not selected by a particular joint conditional analysis, but identified by either of the other two analyses, summary statistics were calculated for association analysis of the SNP conditioned on the GCTA-selected SNP(s).

Genetic risk score. We calculated a genetic risk score for each individual in the population-based KORA study (1,670 men and 1,750 women) using the 49 identified variants, weighted by the allelic effect from the European ancestry meta-analyses of WHRadjBMI. Sex-combined scores were computed on the basis of the sex-combined meta-analysis. Sex-stratified scores were calculated on the basis of men- and women-specific meta-analyses, in which SNPs not achieving nominal significance in the respective sex ($P \geq 0.05$) were excluded. For each individual, the sex-combined and sex-stratified risk scores were rounded to the nearest integer for plotting. Risk scores were then tested for association with WHRadjBMI using linear regression.

Explained variance. We calculated the variance explained by a single SNP as:

$$2 \cdot \text{MAF} \cdot (1 - \text{MAF}) \cdot \frac{\beta^2}{\text{Var}(Y)}$$

in which MAF is the minor allele frequency, β is the SNP effect estimate computed by meta-analysis, and $\text{Var}(Y)$ is the variance of the phenotype Y as it went into the study-specific association testing. To derive the total variance explained by a set of

independent SNPs, we computed the sum of single-SNP explained variances under the assumption of independent contributions.

To estimate the polygenic variance explained by all HapMap SNPs, we used the all-SNP estimation approach implemented in GCTA and analysed individuals in the ARIC and TwinGene cohorts, including the first 20 principal components as fixed covariates. After removing one of each pair of individuals with estimated genetic relatedness >0.025 , 11,898 unrelated individuals with WHRadjBMI were available.

Fine-mapping analyses. We considered each identified locus, defined as 500 kb upstream and downstream of the lead SNP, and computed 99% credible intervals using a Bayesian approach. On the basis of association summary statistics from the European ancestry, non-European ancestry, or all ancestries sex-combined meta-analyses, we calculated an approximate Bayes' factor⁵⁶ in favour of association, given by:

$$BF_j = \frac{\sqrt{1-R_j}}{\exp\left(-\frac{R_j\beta_j^2}{2\sigma_j^2}\right)}$$

in which β_j is the allelic effect of the j th SNP, with corresponding standard error σ_j , and $R_j = 0.04/(\sigma_j^2 + 0.04)$, which incorporates a $N(0, 0.2^2)$ prior for β_j . This prior gives high probability to small effect sizes, and only small probability to large effect sizes. We then calculated the posterior probability that the j th SNP is causal by:

$$\phi_j = \frac{BF_j}{\sum_k BF_k}$$

in which the summation in the denominator is over all SNPs passing quality control across the locus. We compared the meta-analysis results and credible sets of SNPs likely to contain the causal variant as described⁵⁷. Assuming a single causal variant at each locus, a 99% credible set of variants was then constructed by: (1) ranking all SNPs according to their Bayes' factor; and (2) combining ranked SNPs until their cumulative posterior probability exceeded 0.99. For each locus, we calculated the number of SNPs contained within the 99% credible sets, and the length of the genomic interval covered by these SNPs.

Comparison of loci across traits. To determine whether the identified loci were also associated with any of 22 cardio-metabolic traits, we obtained association data from meta-analysis consortia DIAGRAM (T2D)⁵⁸, CARDIoGRAM-C4D (CAD)⁵⁹, ICBP (diastolic and systolic blood pressure)⁶⁰, GIANT (BMI, height)^{36,37}, GLGC (HDL, LDL, and triglycerides)⁶¹, MAGIC (fasting glucose, fasting insulin, fasting insulin adjusted for BMI, and two-hour glucose)⁶²⁻⁶⁴, ADIPOGen (BMI-adjusted adiponectin)⁶⁵, CKDgen (urine albumin-to-creatinine ratio (UACR), estimated glomerular filtration rate (eGFR), and overall CKD)^{66,67}, ReproGen (age at menarche, age at menopause)^{68,69}, and GEFOS (bone mineral density)⁷⁰; others provided association data for IgA nephropathy⁷¹ (also K.K., M.C., R.P.L. and A.G.G., unpublished data) and for endometriosis (stage B cases only)⁷². Proxies ($r^2 > 0.8$ in CEU) were used when an index SNP was unavailable.

We also searched the NHGRI GWAS catalogue for previous SNP-trait associations near our lead SNPs⁷³. We supplemented the catalogue with additional genome-wide significant SNP-trait associations from the literature^{13,70,74-80}. We used PLINK to identify SNPs within 500 kb of lead SNPs using 1000 Genomes Project pilot I genotype data and LD (r^2) values from CEU^{81,82}; for rs7759742, HapMap release 22 CEU data^{81,83} were used. All SNPs within the specified regions were compared with the NHGRI GWAS catalogue¹⁶.

Enrichment of concordant cross-trait associations and effects. To evaluate whether the alleles associated with increased WHRadjBMI at the 49 identified SNPs convey effects for any of the 22 cardiometabolic traits, we conducted meta-regression analyses of the beta-estimates on these metabolic outcomes from other consortia with the beta-estimates for WHRadjBMI in our data⁶⁵.

On the basis of the association data across traits, we generated a matrix of Z-scores by dividing the association betas for each of the 49 WHRadjBMI SNPs for each of 22 traits by their respective standard errors. The traits did not include WHRadjBMI or nephropathy in Chinese subjects, but did include HIPadjBMI and WCadjBMI. Each Z-score was made positive if the original trait-increasing allele also increased the look-up trait and negative if not. Missing associations with were assigned a value of zero. We performed unsupervised hierarchical clustering of the Z score matrix in R using the default settings of the 'heatmap' function from the made4 library (version 1.20.0), agglomerating clusters using average linkage and Pearson correlation metric distance. The rows and columns of matrix values were each automatically scaled to range from 3 to -3. Confidence in the hierarchical clustering was assessed by bootstrap analysis (10,000 resamplings) using the R package 'pvclust'⁸⁴.

Identification of candidate functional variants. The 1000 Genomes CEU pilot data were queried for SNPs within 500 kb and in LD ($r^2 > 0.7$, an arbitrary threshold) with any index SNP. All identified variants were then annotated based on RefSeq transcripts using Annovar to identify potential nonsynonymous variants near identified association signals. The distance between each variant and the nearest transcription start site were calculated using gene annotations from GENCODE (v.12).

To investigate whether SNPs in LD with index SNPs are also in LD with common copy number variants (CNVs), we extracted waist trait association results for a list of SNP proxies that are in high LD ($r^2 > 0.8$, CEU) with CNVs in European populations as described previously⁷. Altogether 6,200 CNV-tagging SNPs were used, which are estimated collectively to capture $>40\%$ of CNVs >1 kb in size.

eQTLs. We examined our lead SNPs in eQTL data sets from several sources (Supplementary Note) for *cis* effects significant at $P < 10^{-5}$. We then checked if the trait-associated SNP also had the strongest association with the expression level of its corresponding transcript. If not, we identified a nearby SNP that had a stronger association with expression (peak transcript SNP) of that transcript. To check whether effects of the peak transcript SNP and waist trait-associated SNP overlapped, we conducted conditional analyses to estimate associations between the waist-associated SNP and transcript level when the peak-transcript-associated SNP was also included in the model, and vice versa. If the association for the expression-associated SNP was not significant ($P > 0.05$) when conditioned on the waist-associated SNP, we concluded that the waist-associated SNP is likely to explain a substantial proportion of the variance in gene transcript levels in the region. For SNPs that passed these criteria in either women or men eQTL data sets from deCODE, we investigated sex heterogeneity in gene transcript levels for whole blood (312 men, 435 women) and subcutaneous adipose tissue (252 men, 351 women) based on the sex-specific beta estimates and standard errors, while accounting for potential correlation between the sex-specific associations⁸.

Epigenomic regulatory element overlap with individual variants. We examined overlap of regulatory elements with the 68 trait-associated variants and variants in LD with them ($r^2 > 0.7$, 1000 Genomes phase 1 version 2 EUR⁸⁵), totalling 1,547 variants. We obtained regulatory element data sets from the ENCODE Consortium²⁴ and Roadmap Epigenomics Project²⁵ corresponding to eight tissues selected based on a current understanding of WHRadjBMI pathways. The 226 regulatory element data sets included experimentally identified regions of open chromatin (DNase-seq, FAIRE-seq), histone modification (H3K4me1, H3K27ac, H3K4me3, H3K9ac and H3K4me2), and transcription factor binding (Supplementary Table 17). When available, we downloaded data processed during the ENCODE Integrative Analysis²⁴. We processed Roadmap Epigenomics sequencing data with multiple biological replicates using MACS2 (ref. 86) and the same Irreproducible Discovery Rate pipeline used in the ENCODE Integrative Analysis. Roadmap Epigenomics data with only a single replicate was processed using MACS2 alone.

Global enrichment of WHRadjBMI-associated loci in epigenomic data sets. We performed permutation-based tests in a subset of 60 open chromatin (DNase-seq) and histone modification (H3K27ac, H3K4me1, H3K4me3 and H3K9ac) data sets to identify global enrichment of the WHRadjBMI-associated loci. We matched the index SNP at each locus with 500 variants having no evidence of association ($P > 0.5$, ~ 1.2 million total variants) with a similar distance to the nearest gene ($\pm 11,655$ bp), number of variants in LD (± 8 variants), and minor allele frequency. Using these pools, we created 10,000 sets of control variants for each of the 49 loci and identified variants in LD ($r^2 > 0.7$) and within 1 Mb. For each SNP set, we calculated the number of loci with at least one variant located in a regulatory region under the assumption that one regulatory variant is responsible for each association signal. We initially calculated an enrichment P value by finding the proportion of control sets for which as many or more loci overlap a regulatory element than the set of associated loci. For increased P value accuracy, we estimated the P value assuming a sum of binomial distributions to represent the number of index SNPs or their LD proxies that overlap a regulatory data set compared to the 500 matched control sets.

GRAIL. Gene Relationships Among Implicated Loci (GRAIL)¹⁹ is a text-mining algorithm that evaluates the degree of relatedness among genes within trait regions. Using PubMed abstracts, a subset of genes enriched for relatedness and a set of keywords that suggest putative pathways are identified. To avoid potential bias from papers investigating candidate genes stimulated by GWAS, we restricted our search to abstracts published before 2006. We tested for enrichment of connectivity in the independent SNPs that were significant in our study at $P < 10^{-5}$.

MAGENTA. To investigate whether pathways including predefined sets of genes were enriched in the lower part of the gene P value distribution for WHRadjBMI, we performed a pathway analysis using Magenta 2.4 (ref. 20) and SNPs present in both the MetaboChip and GWAS meta-analyses. SNPs were assigned to a gene if within 110 kb upstream or 40 kb downstream of the transcript's boundaries. The most significant SNP P value within this interval was adjusted for putative confounders (gene size, number of SNPs in a gene, LD pattern) using stepwise linear

regression, creating a gene association score. If the same SNP was assigned to multiple genes, only the gene with the lowest gene score was kept. The *HLA* region was removed from further analyses owing to its high LD structure and gene density. Each gene was then assigned pathway terms using Gene Ontology (GO), PANTHER, Ingenuity and Kyoto Encyclopedia of Genes and Genomes (KEGG)^{87–90}. Finally, the genes were ranked based on their gene association score, and a modified gene-set enrichment analysis using MAGENTA was performed. This analysis tested for enrichment of gene association score ranks above a given rank cutoff (including 5% of all genes) in a gene-set belonging to a predefined pathway term, compared to multiple, equally sized gene-sets that were randomly sampled from all genes in the genome. Around 10,000–1,000,000 gene-set permutations were performed. DEPICT. This method is described in detail elsewhere^{23,36}. In brief, DEPICT uses gene expression data derived from a panel of 77,840 expression arrays⁹¹, 5,984 molecular pathways (based on 169,810 high-confidence experimentally derived protein–protein interactions⁹²), 2,473 phenotypic gene sets (based on 211,882 gene-phenotype pairs from the Mouse Genetics Initiative⁹³), 737 reactome pathways⁹⁴, 184 KEGG pathways⁹⁵, and 5,083 GO terms¹⁹. DEPICT uses the expression data to reconstitute the protein–protein interaction gene sets, mouse phenotype gene sets, reactome pathway gene sets, KEGG pathway gene sets, and GO term gene sets. To avoid biasing the identification of genes and pathways covered by SNPs on the Metabochip, analyses were restricted to GWAS cohort data and included 226 WHRadjBMI SNPs in 78 non-overlapping loci with sex-combined $P < 10^{-5}$. We used DEPICT to map genes to associated WHRadjBMI loci, which then allowed us to (1) systematically identify the most likely causal gene(s) in a given associated region, (2) identify reconstituted gene sets that were enriched in genes from associated regions, and (3) identify tissue and cell type annotations in which genes from associated regions were highly expressed. Associated regions were defined by all genes residing within LD ($r^2 > 0.5$) distance of the WHRadjBMI-associated index SNPs. Overlapping regions were merged, and SNPs that mapped near to or within the *HLA* region were excluded. The 93 WHRadjBMI SNPs with $P < 10^{-5}$ (clumping thresholds: HapMap release 27 CEU $r^2 = 0.01$, 500 kb) resulted in 78 non-overlapping regions. GWAS+Metabochip index SNPs were annotated with DEPICT-prioritized genes if the DEPICT (GWAS-only) SNP was located within 500 kb. To mark related gene sets, we first quantified significant gene sets' pairwise overlap using a non-probabilistic version of the reconstituted gene sets and the Jaccard index measure. Groups of gene sets with mutual Jaccard indices > 0.25 were subsequently referred to as meta gene sets and named by the most significant gene set in the group (Supplementary Table 18 and Fig. 2a). In Fig. 2a, b, gene sets with similarities between 0.1 and 0.25 were connected by an edge that was scaled according to degree of similarity. The Cytoscape tool was used to construct parts of Fig. 2 (ref. 96). In Fig. 2c, we show the significance of all cell type annotations and annotations that were categorized as 'tissues' at the outermost level of the Medical Subject Heading ontology.

44. Winkler, T. W. *et al.* Quality control and conduct of genome-wide association meta-analyses. *Nature Protocols* **9**, 1192–1212 (2014).

45. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55**, 997–1004 (1999).

46. Buyske, S. *et al.* Evaluation of the metabochip genotyping array in African Americans and implications for fine mapping of GWAS-identified loci: the PAGE study. *PLoS ONE* **7**, e35651 (2012).

47. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).

48. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Series B Stat. Methodol.* **57**, 289–300 (1995).

49. Higgins, J. P. & Thompson, S. G. Quantifying heterogeneity in a meta-analysis. *Stat. Med.* **21**, 1539–1558 (2002).

50. Neale, M. C., Cardon, L. R. & North Atlantic Treaty Organization. *Methodology for Genetic Studies of Twins and Families* (Kluwer Academic Publishers, 1992).

51. Falconer, D. S. *Introduction to Quantitative Genetics* 3rd edn (Oliver and Boyd, 1990).

52. Almasy, L. & Blangero, J. Multipoint quantitative-trait linkage analysis in general pedigrees. *Am. J. Hum. Genet.* **62**, 1198–1211 (1998).

53. Neale, M. C. *MX: Statistical Modeling* 4th edn (Department of Psychiatry, 1997).

54. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).

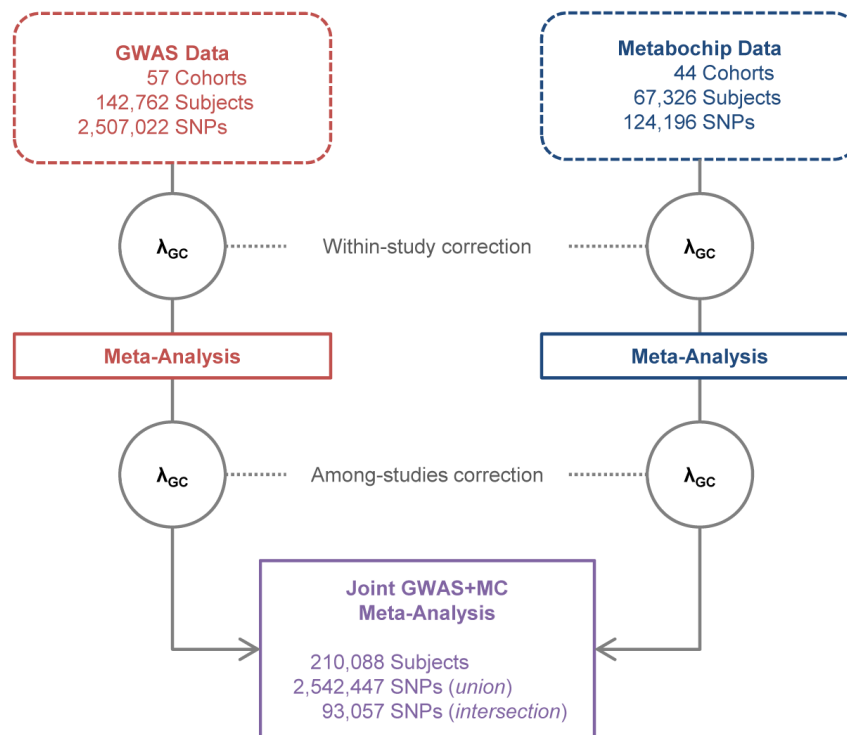
55. Frazer, K. A. *et al.* A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449**, 851–861 (2007).

56. Wakefield, J. A Bayesian measure of the probability of false discovery in genetic epidemiology studies. *Am. J. Hum. Genet.* **81**, 208–227 (2007).

57. Wellcome Trust Case Control Consortium. Bayesian refinement of association signals for 14 loci in 3 common diseases. *Nature Genet.* **44**, 1294–1301 (2012).

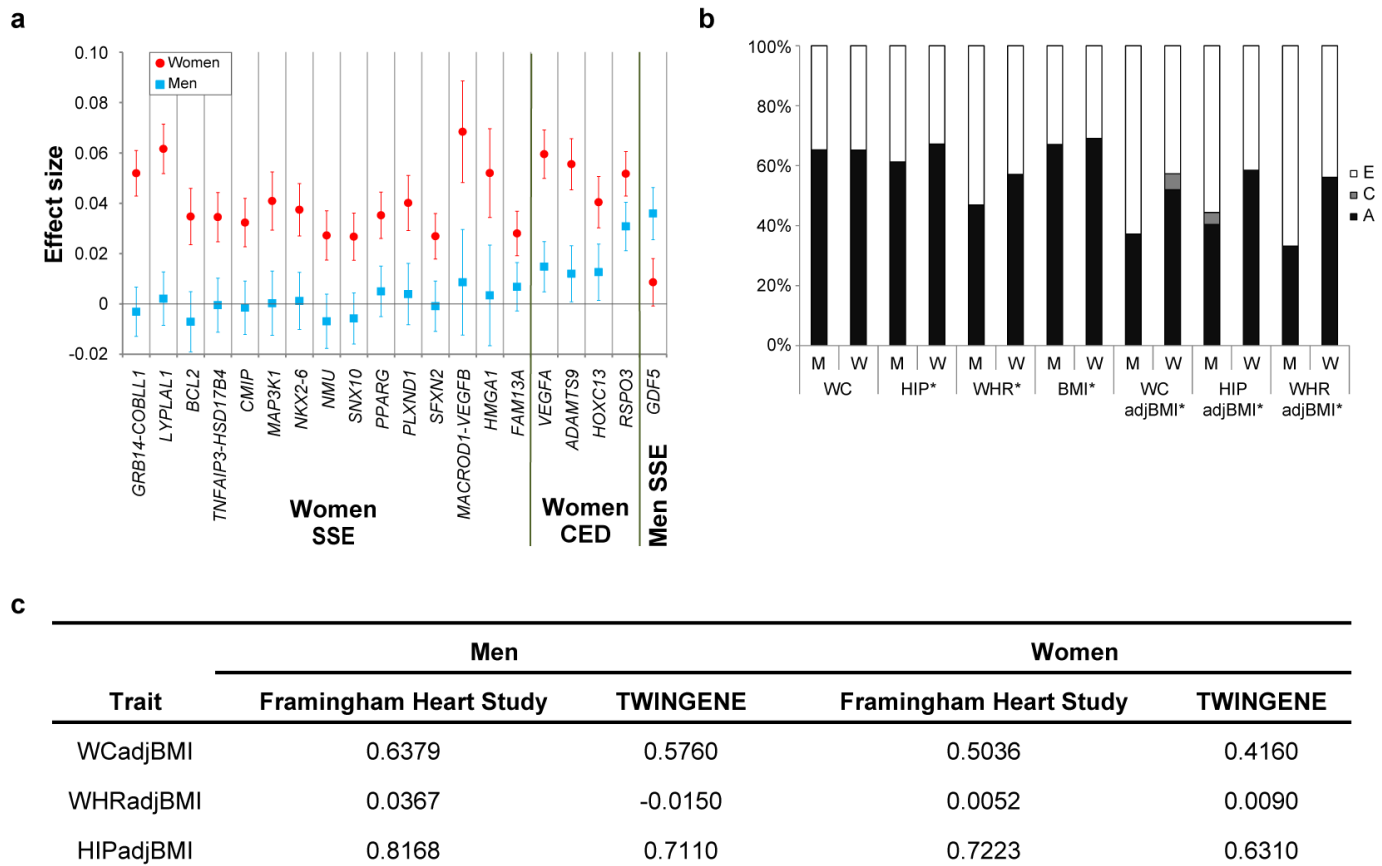
58. Morris, A. P. *et al.* Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nature Genet.* **44**, 981–990 (2012).

59. Deloukas, P. *et al.* Large-scale association analysis identifies new risk loci for coronary artery disease. *Nature Genet.* **45**, 25–33 (2013).
60. Ehret, G. B. *et al.* Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature* **478**, 103–109 (2011).
61. Global Lipids Genetics Consortium. Discovery and refinement of loci associated with lipid levels. *Nature Genet.* **45**, 1274–1283 (2013).
62. Scott, R. A. *et al.* Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. *Nature Genet.* **44**, 991–1005 (2012).
63. Manning, A. K. *et al.* A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. *Nature Genet.* **44**, 659–669 (2012).
64. Saxena, R. *et al.* Genetic variation in *GIPR* influences the glucose and insulin responses to an oral glucose challenge. *Nature Genet.* **42**, 142–148 (2010).
65. Dastani, Z. *et al.* Novel loci for adiponectin levels and their influence on type 2 diabetes and metabolic traits: a multi-ethnic meta-analysis of 45,891 individuals. *PLoS Genet.* **8**, e1002607 (2012).
66. Pattaro, C. *et al.* Genome-wide association and functional follow-up reveals new loci for kidney function. *PLoS Genet.* **8**, e1002584 (2012).
67. Böger, C. A. *et al.* *CUBN* is a gene locus for albuminuria. *J. Am. Soc. Nephrol.* **22**, 555–570 (2011).
68. Stolk, L. *et al.* Meta-analyses identify 13 loci associated with age at menopause and highlight DNA repair and immune pathways. *Nature Genet.* **44**, 260–268 (2012).
69. Elks, C. E. *et al.* Thirty new loci for age at menarche identified by a meta-analysis of genome-wide association studies. *Nature Genet.* **42**, 1077–1085 (2010).
70. Estrada, K. *et al.* Genome-wide meta-analysis identifies 56 bone mineral density loci and reveals 14 loci associated with risk of fracture. *Nature Genet.* **44**, 491–501 (2012).
71. Gharavi, A. G. *et al.* Genome-wide association study identifies susceptibility loci for IgA nephropathy. *Nature Genet.* **43**, 321–327 (2011).
72. Painter, J. N. *et al.* Genome-wide association study identifies a locus at 7p15.2 associated with endometriosis. *Nature Genet.* **43**, 51–54 (2011).
73. Hindorf, L. A. *et al.* Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl Acad. Sci. USA* **106**, 9362–9367 (2009).
74. Kamatani, Y. *et al.* Genome-wide association study of hematological and biochemical traits in a Japanese population. *Nature Genet.* **42**, 210–215 (2010).
75. Franke, A. *et al.* Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nature Genet.* **42**, 1118–1125 (2010).
76. Sawcer, S. *et al.* Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature* **476**, 214–219 (2011).
77. Wang, K. S., Liu, X. F. & Aragam, N. A genome-wide meta-analysis identifies novel loci associated with schizophrenia and bipolar disorder. *Schizophr. Res.* **124**, 192–199 (2010).
78. Cirulli, E. T. *et al.* Common genetic variation and performance on standardized cognitive tests. *Eur. J. Hum. Genet.* **18**, 815–820 (2010).
79. Gieger, C. *et al.* New gene functions in megakaryopoiesis and platelet formation. *Nature* **480**, 201–208 (2011).
80. Need, A. C. *et al.* A genome-wide study of common SNPs and CNVs in cognitive performance in the CANTAB. *Hum. Mol. Genet.* **18**, 4650–4661 (2009).
81. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
82. The 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–1073 (2010).
83. The International HapMap Project. *Nature* **426**, 789–796 (2003).
84. Suzuki, R. & Shimodaira, H. Pvcust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics* **22**, 1540–1542 (2006).
85. 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012).
86. Feng, J., Liu, T., Qin, B., Zhang, Y. & Liu, X. S. Identifying ChIP-seq enrichment using MACS. *Nature Protocols* **7**, 1728–1740 (2012).
87. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature Genet.* **25**, 25–29 (2000).
88. Mi, H. & Thomas, P. PANTHER pathway: an ontology-based pathway database coupled with data analysis tools. *Methods Mol. Biol.* **563**, 123–140 (2009).
89. Jimenez-Marin, A., Collado-Romero, M., Ramirez-Boo, M., Arce, C. & Garrido, J. J. Biological pathway analysis by ArrayUnlock and Ingenuity Pathway Analysis. *BMC Proc.* **3**, S6 (2009).
90. Kanehisa, M. & Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
91. Fehrmann, R. S. *et al.* Gene expression analysis identified global gene dosage sensitivity in cancer. *Nature Genet.* **47**, 115–125 (2015).
92. Lage, K. *et al.* A human phenome-interactome network of protein complexes implicated in genetic disorders. *Nature Biotechnol.* **25**, 309–316 (2007).
93. Bult, C. J. *et al.* Mouse genome informatics in a new age of biological inquiry. *IEEE Int. Symposium Bio-Informatics Biomedical Engineering* 29–32 (2000).
94. Croft, D. *et al.* Reactome: a database of reactions, pathways and biological processes. *Nucleic Acids Res.* **39**, D691–D697 (2011).
95. Kanehisa, M., Goto, S., Sato, Y., Furumichi, M. & Tanabe, M. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res.* **40**, D109–D114 (2012).
96. Saito, R. *et al.* A travel guide to Cytoscape plugins. *Nature Methods* **9**, 1069–1076 (2012).



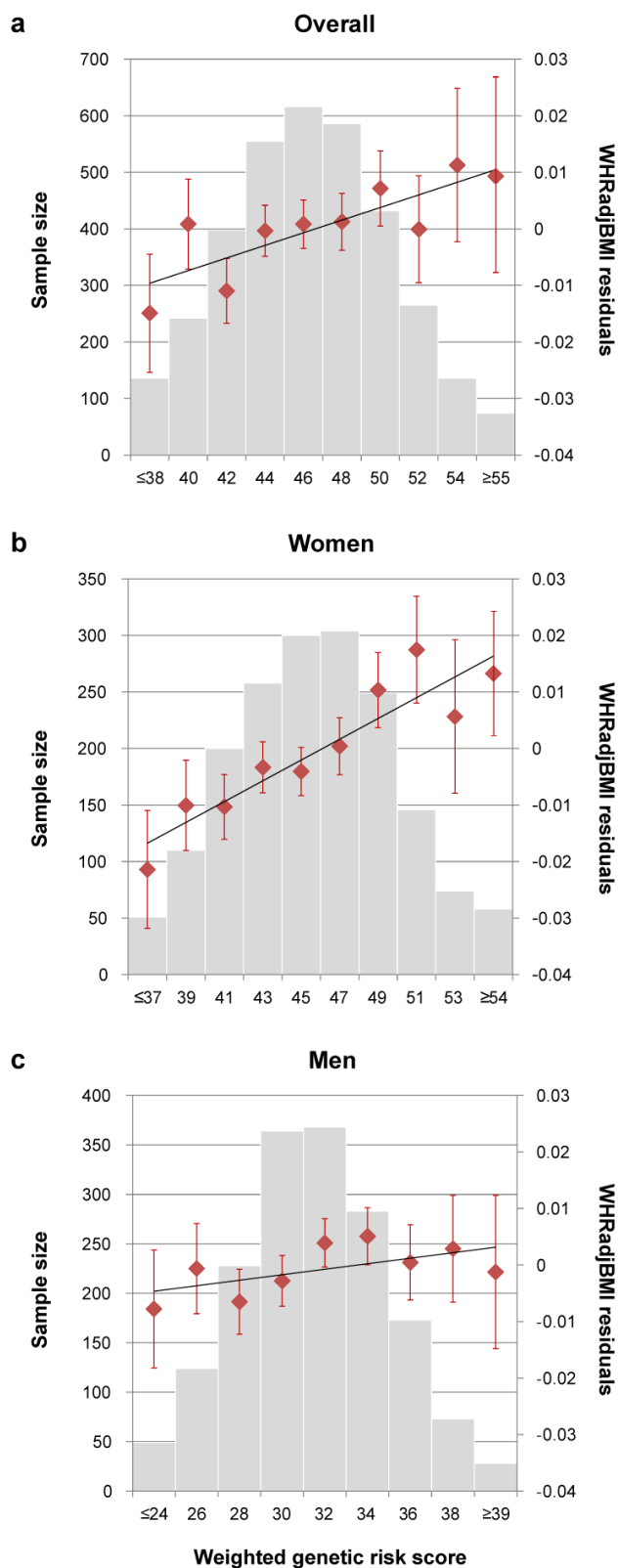
Extended Data Figure 1 | Overall WHRadjBMI meta-analysis study design. Data (dashed lines) and analyses (solid lines) related to the GWAS cohorts for WHRadjBMI are coloured red and those related to the Metabohip (MC) cohorts are coloured blue. The two genomic control (λ_{GC}) corrections (within-study and among-studies) performed on associations from each data set are represented by grey-outlined circles. The λ_{GC} corrections for the GWAS meta-analysis were based on all SNPs and the λ_{GC} corrections for the Metabohip meta-analysis were based on a null set of 4,319 SNPs previously associated with QT interval. The joint meta-analysis of the GWAS and

Metabohip data sets is coloured purple. All SNP counts reflect a sample size filter of $n \geq 50,000$ subjects. Additional WHRadjBMI meta-analyses included Metabohip data from up to 14,371 subjects of east Asian, south Asian or African-American ancestry from eight cohorts. Counts for the meta-analyses of waist circumference, hip circumference, and their BMI-adjusted counterparts (WCadjBMI and HIPadjBMI) differ from those of WHRadjBMI because some cohorts only had phenotype data available for one type of body circumference measurement (see Supplementary Table 2).

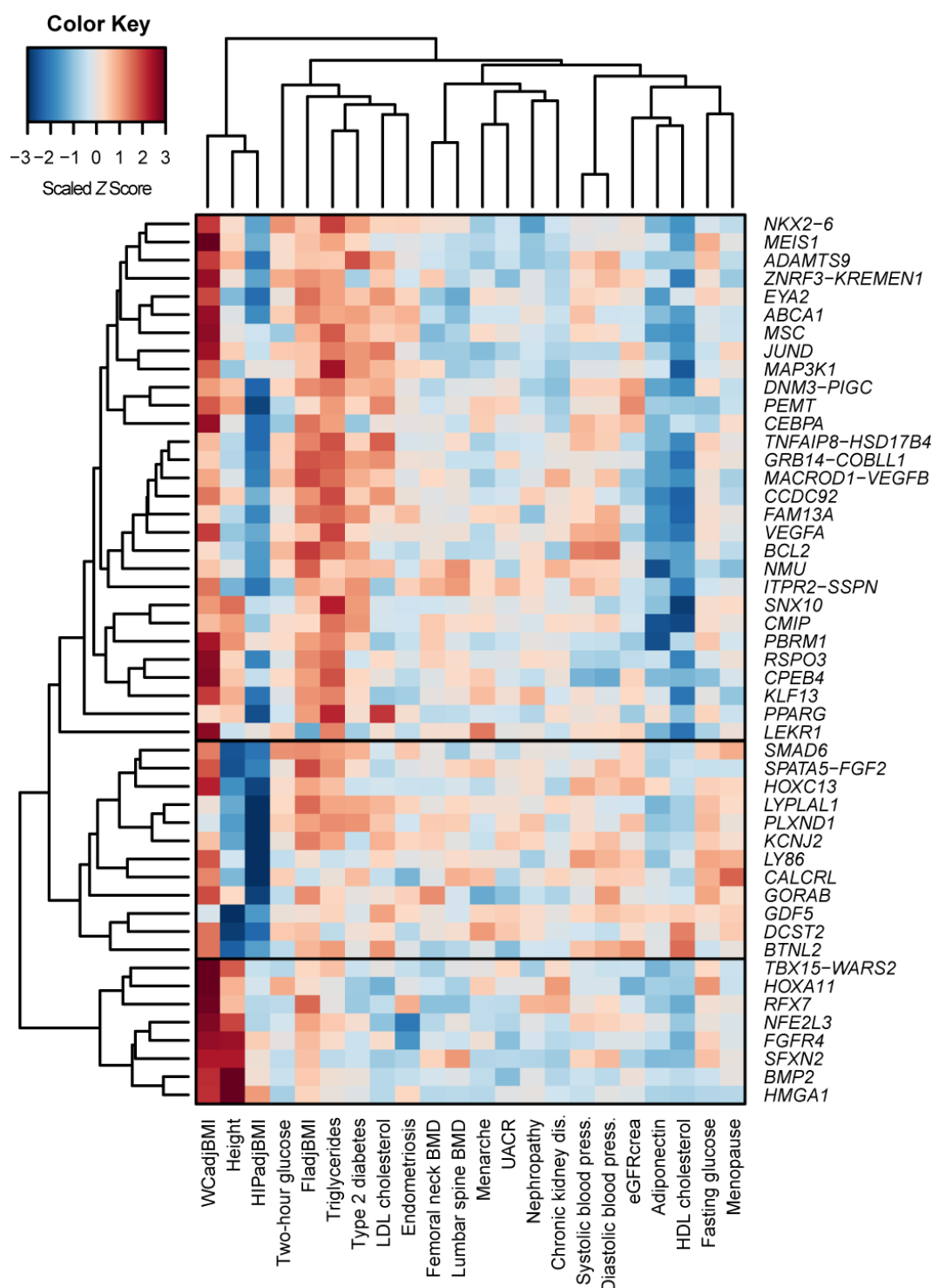


Extended Data Figure 2 | Women- and men-specific effects, phenotypic variances and genetic correlations. **a**, Figure showing effect beta estimates for the 20 WHRadjBMI SNPs showing significant evidence of sexual dimorphism. Sex-specific effect betas and 95% confidence intervals for SNPs associated with WHRadjBMI are shown as red circles and blue squares for women and men, respectively. Sample sizes, comprising more than 73,576 men and 96,182 women, are listed in Table 1. The SNPs are classified into three categories: (1) those showing a women-specific effect ('women SSE'), namely a significant effect in women and no effect in men ($P_{\text{women}} < 5 \times 10^{-8}$, $P_{\text{men}} \geq 0.05$), (2) those showing a pronounced women effect ('women CED'), namely a significant effect in women and a less significant but directionally consistent effect in men ($P_{\text{women}} < 5 \times 10^{-8}$, $5 \times 10^{-8} < P_{\text{men}} \leq 0.05$); and (3) those showing a men-specific effect ('men SSE'), namely a significant effect in men and no effect in women ($P_{\text{men}} < 5 \times 10^{-8}$, $P_{\text{women}} \geq 0.05$). Within each of the

three categories, the loci were sorted by increasing P value of sex-based heterogeneity in the effect betas. **b**, Figure showing standardized sex-specific phenotypic variance components for six waist-related traits. Values are shown in men (M) and women (W) from the Swedish Twin Registry ($n = 11,875$). The ACE models are decomposed into additive genetic components (A) shown in black, common environmental components (C) in grey, and non-shared environmental components (E) in white. Components are shown for waist circumference (WC), hip circumference (HIP), WHR, WCadjBMI, HIPadjBMI and WHRadjBMI. When the 'A' component is different in men and women with $P < 0.05$ for a given trait, its name is marked with an asterisk. **c**, Genetic correlations of waist-related traits with height, adjusted for age and BMI. Genetic correlations of three traits with height were based on variance component models in the Framingham Heart Study and TwinGene study (see Methods).



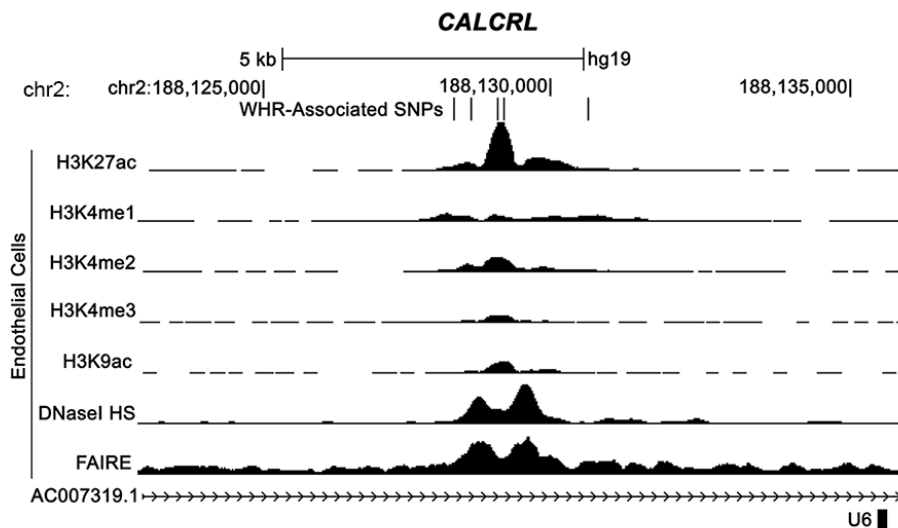
Extended Data Figure 3 | Cumulative genetic risk scores for WHRadjBMI applied to the KORA study cohort. **a**, All subjects ($n = 3,440$, $P_{\text{trend}} = 6.7 \times 10^{-4}$). **b**, Only women ($n = 1,750$, $P_{\text{trend}} = 1.0 \times 10^{-11}$). **c**, Only men ($n = 1,690$, $P_{\text{trend}} = 0.02$). Each genetic risk score illustrates the joint effect of the WHRadjBMI-increasing alleles of the 49 identified variants from Table 1 weighted by the relative effect sizes from the applicable sex-combined or sex-specific meta-analysis. The mean WHRadjBMI residual and 95% confidence interval is plotted for each genetic risk score category (red dots). The histograms show each genetic risk score is normally distributed in KORA (grey bars).



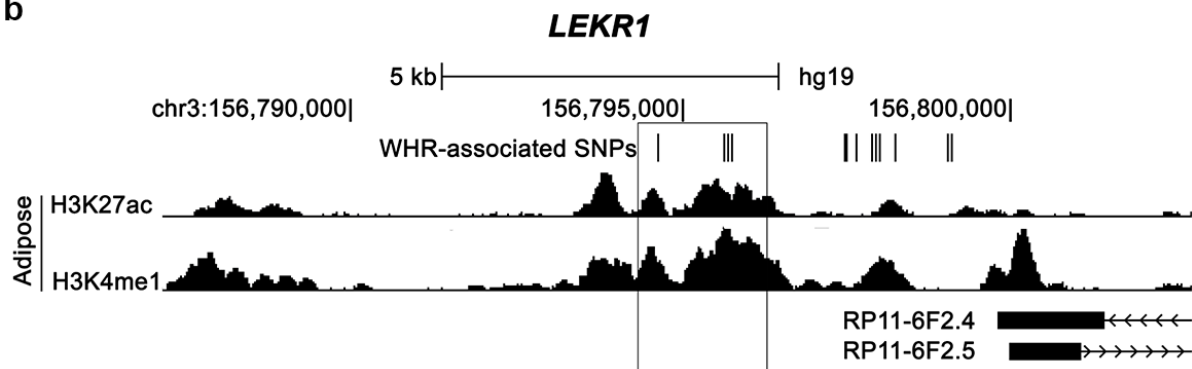
Extended Data Figure 4 | Heat map of unsupervised hierarchical clustering of the effects of 49 WHRadjBMI SNPs on 22 anthropometric and metabolic traits and diseases. The matrix of Z-scores representing the set of associations was scaled by row (locus name) and by column (trait) to range from -3 to 3 . Negative values (blue) indicate that the WHRadjBMI-increasing allele was associated with decreased values of the trait and positive values (red) indicate that this allele was associated with increased values of the trait. Sample sizes for the associations are listed in Supplementary Table 8. Dendrograms indicating the clustering relationships are shown to the left and above the heat map. The WHRadjBMI-increasing alleles at the 49 lead SNPs

segregate into three major clusters comprised of alleles that associate with: (1) larger WCadjBMI and smaller HIPadjBMI ($n = 30$ SNPs); (2) taller stature and larger WCadjBMI ($n = 8$ SNPs); and (3) shorter stature and smaller HIPadjBMI ($n = 11$ SNPs). The three visually identified SNP clusters could be statistically distinguished with $>90\%$ confidence. Alleles of the first cluster were predominantly associated with lower high density lipoprotein (HDL) cholesterol and with higher triglycerides and fasting insulin adjusted for BMI (F1adjBMI). BMD, bone mineral density; eGFRcrea, estimated glomerular filtration rate based on creatinine; LDL cholesterol, low-density lipoprotein cholesterol; UACR, urine albumin-to-creatinine ratio.

a



b



Extended Data Figure 5 | Regulatory element overlap with WHRadjBMI-associated loci. **a**, Five variants associated with WHRadjBMI and located ~77 kb upstream of the first *CALCRL* transcription start site overlap regions with genomic evidence of regulatory activity in endothelial cells. **b**, Five WHRadjBMI variants, including rs8817452, in a 1.1-kb region (box) ~250 kb

downstream of the first *LEKR1* transcription start site overlap evidence of active enhancer activity in adipose nuclei. Signal enrichment tracks are from the ENCODE Integrative Analysis and the Roadmap Epigenomics track hubs on the UCSC Genome Browser. Transcripts are from the GENCODE basic annotation.

Extended Data Table 1 | WHRadjBMI loci with multiple association signals in the sex-combined and/or sex-specific approximate conditional meta-analyses

Locus [*]	SNP	Position (bp)	Nearest gene(s)	EA [†]	EAF	Sex-combined			Women			Men			Sex diff. P^{\ddagger}	CEU r^2 with lead SNP
						β	P	N	β	P	N	β	P	N		
<i>TBX15</i> -	rs2645294	119,376,110	<i>WARS2</i>	T	0.6	0.031	7.60E-19	209,808	0.035	1.50E-14	116,596	0.014	2.20E-02	93,346	4.90E-03	Same
<i>WARS2</i>	rs1106529	119,333,020	<i>TBX15</i>	A	0.8	0.016	1.40E-03	209,930	0.021	1.10E-03	116,663	0.034	4.80E-09	93,401	1.10E-01	0.43
[chr 1]	rs12143789	119,298,677	<i>TBX15</i>	C	0.2	0.026	1.00E-09	209,874	0.022	1.30E-04	116,640	0.019	2.30E-03	93,369	7.10E-01	0.06
	rs12731372	118,654,498	<i>SPAG17</i>	C	0.8	0.024	1.30E-09	209,856	0.02	1.10E-04	116,636	0.028	3.40E-06	93,354	2.80E-01	>500 kb
<i>GRB14</i> -	rs1128249	165,236,870	<i>COBLL1</i>	G	0.6	0.062	8.60E-19	209,414	0.093	1.00E-24	116,348	-0.002	7.10E-01	93,200	8.60E-22	0.93
<i>COBLL1</i>	rs12692737	165,262,555	<i>COBLL1</i>	A	0.3	0.043	1.60E-08	203,265	0.134	2.70E-26	112,317	0.003	5.70E-01	91,082	2.80E-21	0.71
[chr 2]	rs12692738	165,266,498	<i>COBLL1</i>	T	0.8	0.021	5.90E-05	209,551	0.092	3.80E-20	116,474	-0.005	4.10E-01	93,211	4.70E-18	0.3
	rs17185198	165,268,482	<i>COBLL1</i>	A	0.8	0.002	7.40E-01	207,702	0.072	8.50E-13	115,657	-0.004	5.80E-01	92,179	8.00E-11	0.15
<i>PRBM1</i>	rs13083798	52,624,788	<i>PRBM1</i>	A	0.5	0.023	4.10E-11	209,128	0.013	1.20E-01	115,974	0.016	1.10E-03	93,288	7.40E-01	0.88
[chr3]	rs12489828	52,542,054	<i>NT5DC2</i>	T	0.6	0.011	6.50E-02	204,485	0.029	2.60E-10	112,633	-0.015	2.90E-03	91,986	7.20E-11	0.57
<i>MAP3K1</i>	rs3936510	55,896,623	<i>MAP3K1</i>	T	0.2	0.022	1.50E-06	207,896	0.042	6.00E-12	115,645	-0.002	8.20E-01	92,386	5.90E-07	0.88
[chr 5]	rs459193	55,842,508	<i>ANKRD55</i>	A	0.3	0.026	1.60E-11	209,952	0.016	1.90E-03	116,677	0.033	6.70E-09	93,410	2.30E-02	0.06
<i>VEGFA</i>	rs998584 [§]	43,865,874	<i>VEGFA</i>	A	0.5	0.043	1.10E-29	189,620	0.065	1.00E-35	106,771	0.018	8.20E-04	82,983	3.10E-10	0.84
[chr 6]	rs4714699	43,910,541	<i>VEGFA</i>	C	0.4	0.019	3.50E-07	193,327	0.028	1.00E-08	107,987	0.007	1.90E-01	85,475	4.90E-03	0.01
<i>RSPO3</i>	rs1936805 [§]	127,493,809	<i>RSPO3</i>	T	0.5	0.038	2.00E-28	209,859	0.071	6.40E-37	116,602	0.031	3.30E-10	93,392	8.40E-08	Same
[chr 6]	rs11961815	127,477,288	<i>RSPO3</i>	A	0.8	0.022	5.00E-06	209,679	0.037	6.50E-09	116,503	0.021	3.60E-03	93,310	6.90E-02	0.32
	rs72959041	127,496,586	<i>RSPO3</i>	A	0.1	0.101	8.70E-15	72,472	-	-	-	-	-	-	-	0.05
<i>NFE2L3</i>	rs1534696	26,363,764	<i>SNX10</i>	C	0.4	0.011	2.00E-03	198,194	0.028	2.00E-08	111,643	-0.007	1.90E-01	86,685	2.20E-07	Same
<i>SNX10</i> [¶]	rs10245353	25,825,139	<i>NFE2L3</i>	A	0.2	0.035	8.40E-16	210,008	0.016	1.30E-01	116,704	0.027	1.40E-05	93,438	3.60E-01	Same
[chr 7]	rs3902751	25,828,164	<i>NFE2L3</i>	A	0.3	0.009	2.00E-01	209,969	0.039	4.20E-14	116,676	0.019	8.40E-04	93,427	7.40E-03	0.608 [¶]
<i>HOXC13</i>	rs1443512	52,628,951	<i>HOXC13</i>	A	0.2	0.016	2.70E-03	209,980	0.04	1.10E-14	116,688	0.012	3.00E-02	93,425	1.80E-04	Same
[chr 12]	rs10783615	52,636,040	<i>HOXC12</i>	G	0.1	0.037	6.70E-14	209,368	0.023	8.50E-03	116,356	0.022	1.80E-03	93,146	9.30E-01	0.59
	rs2071449 [§]	52,714,278	<i>HOXC4/5/6</i>	A	0.4	0.028	5.00E-15	206,953	0.026	4.60E-08	114,259	0.029	3.40E-08	92,829	6.60E-01	0
<i>CCDC92</i>	rs4765219	123,006,063	<i>CCDC92</i>	C	0.7	0.025	6.90E-12	209,807	0.032	2.50E-11	116,592	0.018	5.30E-04	93,350	3.80E-02	Same
[chr 12]	rs863750	123,071,397	<i>ZNF664</i>	T	0.6	0.022	3.90E-10	209,371	0.031	1.60E-11	116,367	0.015	4.00E-03	93,138	1.80E-02	0.02

P values and β coefficients for the association with WHRadjBMI from the joint model in the approximate conditional analysis of combined GWAS and MetaboChIP studies. SNPs selected by conditional analyses as independently associated with WHRadjBMI in a meta-analysis (sex-combined, women- or men-specific) have their respective summary statistics for these analyses marked in black and bold. SNPs not selected by a particular conditional analysis as independently associated are marked in grey and show the association analysis results for the SNP conditioned on the locus SNPs selected by GCTA. Sample sizes are from the unconditioned meta-analysis.

* Locus and lead SNPs are defined by Table 1.

† The effect allele is the WHRadjBMI-increasing allele in the sex-combined analysis.

‡ Test for sex difference in conditional analysis based on the effect correlation estimate from primary analyses; values significant at the table-wise Bonferroni threshold of $0.05/25 = 2 \times 10^{-3}$ are marked in bold.

§ SNPs selected by conditional analysis in the sex-combined analysis; proxies were selected by joint conditional analysis in the women- and/or men-specific analyses.

|| SNP not present in the sex-specific meta-analyses due to sample size filter requiring $n \geq 50,000$; sample size from GCTA.

¶ At *NFE2L3*-*SNX10*, different lead SNPs were identified in the European and all-ancestry analyses but LD is reported with respect to rs10245353.

Extended Data Table 2 | Enrichments of 49 WHRadjBMI signal SNPs with metabolic and anthropometric traits

Trait	Max. sample size	SNPs in concordant direction			SNPs in concordant direction with $P < 0.05$		
		<i>N</i>	Total	<i>P</i>	<i>N</i>	Total	<i>P</i>
Type 2 diabetes (T2D)	86,200	37	49	2.35E-04	16	49	3.56E-14
Fasting glucose (FG)	132,996	35	49	1.90E-03	8	49	2.75E-05
Fasting insulin adjusted for BMI (FIadjBMI)	103,496	45	49	4.11E-10	36	49	4.04E-47
2-hour glucose (G120)	42,853	33	49	1.06E-02	7	49	2.09E-04
Diastolic blood pressure (DBP)	69,760	34	49	4.70E-03	10	49	3.21E-07
Systolic blood pressure (SBP)	69,774	38	49	7.10E-05	6	49	1.36E-03
Body mass index (BMI)	322,120	40	49	4.63E-06	23	49	4.42E-24
Height	253,209	25	49	5.00E-01	14	49	1.10E-11
High-density lipoprotein cholesterol (HDL-C)	187,142	45	49	4.11E-10	24	49	1.22E-25
Low-density lipoprotein cholesterol (LDL-C)	173,067	33	49	1.06E-02	12	49	2.32E-09
Triglycerides (TG)	177,838	46	49	3.49E-11	29	49	6.02E-34
Adiponectin	29,347	41	49	9.82E-07	20	49	1.28E-19
Endometriosis	1,364/7,060	24	45	3.83E-01	4	45	2.58E-02
Nephropathy (in Chinese subjects)	1,194/902	18	43	8.89E-01	0	43	1.00E+00
Nephropathy (in Italian subjects)	1,045/1,340	20	43	7.29E-01	1	43	6.63E-01
Estimated glomerular filtration rate of creatinine (eGFRcrea)	74,354	29	49	1.26E-01	3	49	1.24E-01
Chronic kidney disease (CKD)	74,354	17	49	9.89E-01	2	49	3.47E-01
Urine albumin-to-creatinine ratio (UACR)	31,580	22	49	8.04E-01	2	49	3.47E-01
Menopause	87,802	28	49	1.96E-01	1	49	7.11E-01
Menarche	38,968	23	49	7.16E-01	2	49	3.47E-01
Coronary artery disease (CAD)	191,198	27	48	2.35E-01	9	48	2.64E-06
Femoral neck bone mineral density (FN-BMD)	32,960	25	49	5.00E-01	4	49	3.40E-02
Lumbar spine bone mineral density (LS-BMD)	31,798	28	49	1.96E-01	3	49	1.24E-01

The 49 WHRadjBMI SNPs were tested for association with other traits by GWAS meta-analyses performed by other groups (see Methods). The maximum sample size available is shown overall or separately for cases/controls. *N* indicates the number of the total SNPs for which the WHRadjBMI-increasing allele is associated with the trait in the concordant direction (increased levels, except for HDL-C, adiponectin and BMI). One-sided binomial *P* values test whether this number is greater than expected by chance (null $P = 0.5$ and null $P = 0.025$, respectively). The tests do not account for correlation between WHRadjBMI and the tested traits. *P* values representing significant column-wise enrichment ($P < 0.05/23$ tests) are marked in bold.

Extended Data Table 3 | Enrichment of 49 WHRadjBMI-associated loci in epigenomic data sets

Sample	Tissue	DNase I HS	H3K4me1	H3K27ac	H3K4me3	H3K9ac
Adipose Nuclei	Adipose	-	9.6E-06	1.2E-13	0.0051	0.0010
GM12878	Blood	0.029	0.032	0.32	0.050	0.030
Osteoblasts	Bone	0.082	4.1E-06	1.8E-04	9.9E-04	-
Astrocytes	Brain	0.013	0.0044	0.0077	0.0047	-
Anterior Caudate	Brain	-	2.9E-04	0.026	0.018	0.015
Mid Frontal Lobe	Brain	-	0.029	0.023	0.023	0.036
Substantia Nigra	Brain	-	0.047	-	0.023	0.045
Cerebellum	Brain	0.048	-	-	-	-
Cerebrum Frontal	Brain	0.054	-	-	-	-
Frontal Cortex	Brain	0.022	-	-	-	-
HUVEC	Endothelial	5.0E-05	0.011	0.0011	0.023	0.040
Adult Liver	Liver	-	0.0057	-	0.15	0.29
HepG2	Liver	0.015	7.7E-05	0.023	5.0E-04	0.085
Hepatocytes	Liver	0.59	-	-	-	-
Huh-7	Liver	0.0024	-	-	-	-
Myocyte	Muscle	2.9E-04	1.3E-04	0.0026	0.015	0.0041
PSOAS	Muscle	0.0012	-	-	-	-
Skeletal Muscle	Muscle	-	7.3E-04	7.8E-05	0.0075	0.25
Pancreatic Islet	Pancreatic Islets	0.40	0.68	-	0.37	0.61

Enrichment of WHRadjBMI-associated loci in regulatory elements from selected WHRadjBMI-relevant tissues. *P* values are derived using a sum of binomial distributions (see Methods). *P* values below a Bonferroni-corrected threshold for 60 tests of 8.3×10^{-4} are indicated in bold font. The binomial-based *P* values are similar to *P* values generated from 10,000 permutation tests. Dashes indicate that data sets were not available.

Extended Data Table 4 | Candidate genes at new loci associated with additional waist and hip-related traits

SNP	Trait	Chr	Locus	Expression QTL ($P < 10^{-5}$) [*]	GRAIL ($P < 0.05$) [†]	Literature [‡]	Other GWAS signals [§]	nsSNPs and CNVs ($r^2 > 0.7$)
rs10925060	WCadjBMI	1	OR2W5- NRLP3	-	-	NLRP3	-	-
rs10929925	HIP	2	SOX11	-	SOX11	SOX11	-	-
rs2124969	WCadjBMI	2	ITGB6	PLA2R1 (SAT)	ITGB6	-	Idiopathic membranous nephropathy (PLA2R1, LY75, ITGB6, RBMS1)	-
rs1664789	WCadjBMI	5	ARL15	-	-	ARL15	-	-
rs17472426	WCadjBMI	5	CCNJL	-	-	FABP6	-	-
rs722585	HIPadjBMI	6	GMDS	-	-	-	-	-
rs7739232	HIPadjBMI	6	KLHL31	KLHL31 (SAT)	-	KLHL31-GCLC- ELVOL	-	-
rs1144	WCadjBMI	7	SRPK2	SRPK2 (LCL), MLL5 (Omental)	-	-	-	-
rs13241538	HIPadjBMI	7	KLF14	KLF14 (SAT)	-	KLF14	HDL cholesterol, Triglycerides, Type 2 diabetes: KLF14	-
rs2398893	WHR	9	PTPDC1	-	BARX1	-	-	-
rs7044106	HIPadjBMI	9	C5	-	-	-	-	-
rs11607976	HIP	11	MYEOV	-	CCND1	FGF19-FGF4-FGF3	-	-
rs1784203	WCadjBMI	11	KIAA1731	-	-	-	-	-
rs1394461	WHR	11	CNTN5	-	-	-	-	-
rs319564	WHR	13	GPC6	-	-	GPC6	-	-
rs4985155	HIP	16	PDXDC1	PDXDC1 (SAT)	-	PLA2G10-NTAN1	Femoral neck bone mineral density, Lumbar spine bone mineral density, Plasma phospholipid levels, Metabolic traits, Height: PDXDC1, NTAN1	NTAN1 (S287P), NTAN1 (H283N)
rs2047937	WCadjBMI	16	ZNF423	-	-	ZNF423-CNEP1R1	-	-
rs2034088	HIPadjBMI	17	VPS53	VPS53 (Liver, SAT), FAM101B (Omental, SAT)	-	-	-	-
rs1053593	HIPadjBMI	22	HMGXB4	TOM1 (PBMC), HMGXB4 (Blood, SAT)	-	HMGXB4	-	HMGXB4 (G165V), CNVR8147.1

Candidate genes for loci shown on Table 3 based on secondary analyses or literature review. Further details are provided in other Supplementary Tables and the Supplementary Note. Loci are shown in order of chromosome and position.

* Gene transcript levels associated with SNP genotype (eQTL) in the indicated tissue(s).

† Genes in pathways identified as enriched by GRAIL analysis.

‡ Strongest candidate genes identified based on manual literature review.

§ Traits associated at $P < 5 \times 10^{-8}$ in GWAS lookups or in the GWAS catalogue using the index SNP or a proxy in high LD ($r^2 > 0.7$), and the gene(s) named in those reports.

|| Non-synonymous variants (nsSNPs) and copy number variants (CNVs) with tag SNPs in high LD with index SNP based on a 1000 Genomes CEU reference panel. DEPICT analysis was not performed for loci associated with these traits.

Genetic studies of body mass index yield new insights for obesity biology

A list of authors and their affiliations appears at the end of the paper

Obesity is heritable and predisposes to many diseases. To understand the genetic basis of obesity better, here we conduct a genome-wide association study and Metabochip meta-analysis of body mass index (BMI), a measure commonly used to define obesity and assess adiposity, in up to 339,224 individuals. This analysis identifies 97 BMI-associated loci ($P < 5 \times 10^{-8}$), 56 of which are novel. Five loci demonstrate clear evidence of several independent association signals, and many loci have significant effects on other metabolic phenotypes. The 97 loci account for $\sim 2.7\%$ of BMI variation, and genome-wide estimates suggest that common variation accounts for $> 20\%$ of BMI variation. Pathway analyses provide strong support for a role of the central nervous system in obesity susceptibility and implicate new genes and pathways, including those related to synaptic function, glutamate signalling, insulin secretion/action, energy metabolism, lipid biology and adipogenesis.

Obesity is a worldwide epidemic associated with increased morbidity and mortality that imposes an enormous burden on individual and public health. Around 40–70% of inter-individual variability in BMI, commonly used to assess obesity, has been attributed to genetic factors^{1–3}. At least 77 loci have previously been associated with an obesity measure⁴, 32 loci from our previous meta-analysis of BMI genome-wide association studies (GWAS)⁵. Nevertheless, most of the genetic variability in BMI remains unexplained. Moreover, although analyses of previous genetic association results have suggested intriguing biological processes underlying obesity susceptibility, few specific genes supported these pathways^{5,6}. For the vast majority of loci, the probable causal gene(s) and pathways remain unknown.

To expand the catalogue of BMI susceptibility loci and gain a better understanding of the genes and biological pathways influencing obesity, we performed the largest GWAS meta-analysis for BMI so far. This work doubles the number of individuals contributing GWAS results, incorporates results from $> 100,000$ individuals genotyped with Metabochip⁷, and nearly doubles the number of BMI-associated loci. Comprehensive assessment of meta-analysis results provides several lines of evidence supporting candidate genes at many loci and highlights pathways that reinforce and expand our understanding of biological processes underlying obesity.

Identification of 97 genome-wide significant loci

This BMI meta-analysis included association results for up to 339,224 individuals from 125 studies, 82 with GWAS results ($n = 236,231$) and 43 with results from Metabochip ($n = 103,047$; Extended Data Table 1 and Supplementary Tables 1–3). After regression on age and sex and inverse normal transformation of the residuals, we carried out association analyses with genotypes or imputed genotype dosages. GWAS were meta-analysed together, as were Metabochip studies, followed by a combined GWAS plus Metabochip meta-analysis. In total, we analysed data from 322,154 individuals of European descent and 17,072 individuals of non-European descent (Extended Data Fig. 1).

Our primary meta-analysis of European-descent individuals from GWAS and Metabochip studies ($n = 322,154$) identified 77 loci reaching genome-wide significance (GWS) and separated by at least 500 kilobases (kb) (Table 1, Extended Data Table 2 and Supplementary Figs 1 and 2). We carried out additional analyses to explore the effects of power and heterogeneity. The inclusion of 17,072 non-European-descent individuals (total $n = 339,224$) identified ten more loci, while secondary

analyses identified another ten GWS loci (Table 2, Supplementary Tables 4–8 and Supplementary Figs 3–9). Of the 97 BMI-associated loci, 41 have previously been associated with one or more obesity measure^{5,8–12}. Thus, our current analyses identified 56 novel loci associated with BMI (Tables 1 and 2 and Extended Data Table 2).

Effects of associated loci on BMI

Newly identified loci generally have lower minor allele frequency and/or smaller effect size estimates than previously known loci (Extended Data Fig. 2a, b). On the basis of effect estimates in the discovery data set, which may be inflated owing to winner's curse, the 97 loci account for 2.7% of BMI phenotypic variance (Supplementary Table 4 and Extended Data Fig. 2a, b). We conservatively used only GWS single nucleotide polymorphisms (SNPs) after strict double genomic control correction, which probably over-corrects association statistics given the lack of evidence for population stratification in family-based analyses¹³ (Extended Data Fig. 3 and Extended Data Table 1). Polygene analyses suggest that SNPs with P values well below GWS add significantly to the phenotypic variance explained. For example, 2,346 SNPs selected from conditional and joint multiple-SNP analysis with $P < 5 \times 10^{-3}$ explained $6.6 \pm 1.1\%$ (mean \pm s.e.m.) of variance, compared to $21.6 \pm 2.2\%$ explained by all HapMap3 SNPs (31–54% of heritability; Fig. 1a). Furthermore, of 1,909 independent SNPs (pairwise distance > 500 kb and $r^2 < 0.1$) included on Metabochip for replication of suggestive BMI associations, 1,458 (76.4%) have directionally consistent effects with our previous GWAS meta-analysis⁵ and the non-overlapping samples in the current meta-analysis (Extended Data Fig. 2c). On the basis of the significant excess of these directionally consistent observations (sign test $P = 2.5 \times 10^{-123}$), we estimate $\sim 1,007$ of the 1,909 SNPs represent true BMI associations.

We compared the effects of our 97 BMI-associated SNPs between the sexes, between ethnicities, and across several cross-sections of our data (Supplementary Tables 4–11 and Extended Data Fig. 4). Two previously identified loci, near *SEC16B* ($P = 5.2 \times 10^{-5}$) and *ZFP64* ($P = 9.1 \times 10^{-5}$), showed evidence of heterogeneity between men and women. Both have stronger effects in women (Supplementary Table 10). Two SNPs, near *NEGR1* ($P = 9.1 \times 10^{-5}$) and *PRKDI* ($P = 1.9 \times 10^{-5}$), exhibited significant evidence for heterogeneity of effect between European- and African-descent samples, and one SNP, near *GBE1* ($P = 1.3 \times 10^{-4}$), exhibited evidence for heterogeneity between European and east Asian individuals (Supplementary Table 9). These findings may reflect true

Table 1 | Novel GWS BMI loci in European meta-analysis

SNP	Chr:position	Notable gene(s)*	Alleles	EAF	β	s.e.m.	P value
rs657452	1:49,362,434	AGBL4(N)	A/G	0.394	0.023	0.003	5.48×10^{-13}
rs12286929	11:114,527,614	CADM1(N)	G/A	0.523	0.022	0.003	1.31×10^{-12}
rs7903146	10:114,748,339	TCF7L2(B,N)	C/T	0.713	0.023	0.003	1.11×10^{-11}
rs10132280	14:24,998,019	STXBP6(N)	C/A	0.682	0.023	0.003	1.14×10^{-11}
rs17094222	10:102,385,430	HIF1AN(N)	C/T	0.211	0.025	0.004	5.94×10^{-11}
rs7599312	2:213,121,476	ERBB4(D,N)	G/A	0.724	0.022	0.003	1.17×10^{-10}
rs2365389	3:61,211,502	FHIT(N)	C/T	0.582	0.020	0.003	1.63×10^{-10}
rs2820292	1:200,050,910	NAV1(N)	C/A	0.555	0.020	0.003	1.83×10^{-10}
rs12885454	14:28,806,589	PRKD1(N)	C/A	0.642	0.021	0.003	1.94×10^{-10}
rs16851483	3:142,758,126	RASA2(N)	T/G	0.066	0.048	0.008	3.55×10^{-10}
rs1167827	7:75,001,105	HIP1(B,N); PMS2L3(B,Q); PMS2P5(Q); WBSCR16(Q)	G/A	0.553	0.020	0.003	6.33×10^{-10}
rs758747	16:3,567,359	NLRC3(N)	T/C	0.265	0.023	0.004	7.47×10^{-10}
rs1928295	9:119,418,304	TLR4(B,N)	T/C	0.548	0.019	0.003	7.91×10^{-10}
rs9925964	16:31,037,396	KAT8(N); ZNF646(M,Q); VKORC1(Q); ZNF668(Q); STX1B(D); FBXL19(D)	A/G	0.620	0.019	0.003	8.11×10^{-10}
rs11126666	2:26,782,315	KCNK3(D,N)	A/G	0.283	0.021	0.003	1.33×10^{-9}
rs2650492	16:28,240,912	SBK1(D,N); APOBR(B)	A/G	0.303	0.021	0.004	1.92×10^{-9}
rs6804842	3:25,081,441	RARB(B)	G/A	0.575	0.019	0.003	2.48×10^{-9}
rs4740619	9:15,624,326	C9orf93(C,M,N)	T/C	0.542	0.018	0.003	4.56×10^{-9}
rs13191362	6:162,953,340	PARK2(B,D,N)	A/G	0.879	0.028	0.005	7.34×10^{-9}
rs3736485	15:49,535,902	SCG3(B,D); DMXL2(M,N)	A/G	0.454	0.018	0.003	7.41×10^{-9}
rs17001654	4:77,348,592	NUP54(M); SCARB2(Q,N)	G/C	0.153	0.031	0.005	7.76×10^{-9}
rs11191560	10:104,859,028	NT5C2(N); CYP17A1(B); SFXN2(Q)	C/T	0.089	0.031	0.005	8.45×10^{-9}
rs1528435	2:181,259,207	UBE2E3(N)	T/C	0.631	0.018	0.003	1.20×10^{-8}
rs1000940	17:5,223,976	RABEP1(N)	G/A	0.320	0.019	0.003	1.28×10^{-8}
rs2033529	6:40,456,631	TDRG1(N); LRFN2(D)	G/A	0.293	0.019	0.003	1.39×10^{-8}
rs11583200	1:50,332,407	ELAVL4(B,D,N,Q)	C/T	0.396	0.018	0.003	1.48×10^{-8}
rs9400239	6:109,084,356	FOXO3(B,N); HSS00296402(Q)	C/T	0.688	0.019	0.003	1.61×10^{-8}
rs10733682	9:128,500,735	LMX1B(B,N)	A/G	0.478	0.017	0.003	1.83×10^{-8}
rs11688816	2:62,906,552	EHBP1(B,N)	G/A	0.525	0.017	0.003	1.89×10^{-8}
rs11057405	12:121,347,850	CLIP1(N)	G/A	0.901	0.031	0.006	2.02×10^{-8}
rs11727676	4:145,878,514	HHIP(B,N)	T/C	0.910	0.036	0.006	2.55×10^{-8}
rs3849570	3:81,874,802	GBE1(B,M,N)	A/C	0.359	0.019	0.003	2.60×10^{-8}
rs6477694	9:110,972,163	EPB41L4B(N); C9orf4(D)	C/T	0.365	0.017	0.003	2.67×10^{-8}
rs7899106	10:87,400,884	GRID1(B,N)	G/A	0.052	0.040	0.007	2.96×10^{-8}
rs2176598	11:43,820,854	HSD17B12(B,M,N)	T/C	0.251	0.020	0.004	2.97×10^{-8}
rs2245368	7:76,446,079	PMS2L11(N)	C/T	0.180	0.032	0.006	3.19×10^{-8}
rs17724992	19:18,315,825	GDF15(B); PGPEP1(Q,N)	A/G	0.746	0.019	0.004	3.42×10^{-8}
rs7243357	18:55,034,299	GRP(B,G,N)	T/G	0.812	0.022	0.004	3.86×10^{-8}
rs2033732	8:85,242,264	RALYL(D,N)	C/T	0.747	0.019	0.004	4.89×10^{-8}

GWS is defined as $P < 5 \times 10^{-8}$. SNP positions are reported according to Build 36 and their alleles are coded based on the positive strand. Alleles (effect/other), effect allele frequency (EAF), beta (β), standard error of the mean (s.e.m.) and P values are based on the meta-analysis of GWAS I+II+MetaboChIP association data from the European sex-combined data set.

* Notable genes from biological relevance to obesity (B); copy number variation (C); DEPICT analyses (D); GRAIL results (G); BMI-associated variant is in strong LD ($r^2 \geq 0.7$) with a missense variant in the indicated gene (M); gene nearest to index SNP (N); association and eQTL data converge to affect gene expression (Q).

heterogeneity at these loci, but are most likely due to linkage disequilibrium (LD) differences across ancestries. Effect estimates for 79% of BMI-associated SNPs in African-descent samples ($P = 9.2 \times 10^{-9}$) and 91% in east Asian samples ($P = 1.8 \times 10^{-15}$) showed directional consistency with our European-only analyses. These results suggest that common BMI-associated SNPs have comparable effects across ancestries and between sexes. In additional heterogeneity analyses, we detected an influence of ascertainment at *TCF7L2* (stronger effects in type 2 diabetes case/control studies than in population-based studies); however, we saw no evidence of systematic ascertainment bias at other loci owing to inclusion of case/control studies (Supplementary Tables 10 and 11).

We also took advantage of LD differences across populations to fine-map association signals using Bayesian methods^{14,15}. At 10 of 27 loci fine-mapped for BMI on MetaboChIP, the addition of non-European individuals into the meta-analysis either narrowed the genomic region containing the 99% credible set, or decreased the number of SNPs in the credible set (Supplementary Table 12 and Supplementary Fig. 10). At the *SEC16B* and *FTO* loci, the all ancestries credible set includes a single SNP, although the SNP we highlight at *FTO* (rs1558902) differs from that identified by a recent fine-mapping effort in African-American cohorts¹⁶. Fine-mapping efforts using larger, more diverse study samples and more complete catalogues of variants will help to further narrow association signals.

We examined the combined effects of lead SNPs at the 97 loci in an independent sample of 8,164 European-descent individuals from the Health and Retirement Study¹⁷. We observed an average increase of

0.1 BMI units (kg per m²) per BMI-increasing allele, equivalent to 260–320 g for an individual 160–180 cm in height. There was a 1.8 kg per m² difference in mean BMI between the 145 individuals (1.78%) carrying the most BMI-increasing alleles (>104) and those carrying the mean number of BMI-increasing alleles in the sample (91; Extended Data Fig. 2d), corresponding to a difference of 4.6–5.8 kg for an individual 160–180 cm in height, and a 1.5 kg per m² difference (3.8–4.9 kg difference) in mean BMI between the 95 individuals (1.16%) carrying the least BMI-increasing alleles (<78) and those carrying the mean number. Such differences are medically significant in predisposing to development of metabolic disease¹⁸. For predicting obesity (BMI ≥ 30 kg per m²), adding genetic risk score to a model including age, age squared, sex and four genotype-based principal components slightly, but significantly increases the area under the receiver-operating characteristic curve from 0.576 to 0.601.

Additional associated variants at BMI loci

To identify additional SNPs with independent BMI associations at the 97 established loci, we used genome-wide complex trait analysis (GCTA)¹⁹ to perform approximate joint and conditional association analysis²⁰ using summary statistics from European sex-combined meta-analysis after removing family-based validation studies (TwinGene and QIMR). GCTA confirmed two signals at *MC4R* previously identified using exact conditional analyses⁵, and identified five loci with evidence of independent associations (Table 3): second signals near *LINC01122*, *NLRC3-ADCY9*, *GPRC5B-GP2* and *BDNF*, and a third signal near *MC4R* (rs9944545,

Table 2 | GWS BMI loci from secondary analyses

SNP	Chr:position	Notable gene(s)*	Alleles	EAF	β	s.e.m.	P value	Analysis
Novel loci								
rs9641123	7:93,035,668	<i>CALCR</i> (B,N); <i>hsa-miR-653</i> (Q)	C/G	0.430	0.029	0.005	2.08×10^{-10}	EPB
rs7164727	15:70,881,044	<i>LOC100287559</i> (N); <i>BBS4</i> (B,M,Q)	T/C	0.671	0.019	0.003	3.92×10^{-9}	All
rs492400	2:219,057,996	<i>PLCD4</i> (B,Q); <i>CYP27A1</i> (B); <i>USP37</i> (N); <i>TLL4</i> (M,Q); <i>STK36</i> (B,M); <i>ZNF142</i> (M); <i>RQCD1</i> (Q)	C/T	0.424	0.024	0.004	6.78×10^{-9}	Men
rs2080454	16:47,620,091	<i>CBLN1</i> (N)	C/A	0.413	0.017	0.003	8.60×10^{-9}	All
rs7239883	18:38,401,669	<i>LOC284260</i> (N); <i>RIT2</i> (B,D)	G/A	0.391	0.023	0.004	1.51×10^{-8}	Women
rs2836754	21:39,213,610	<i>ETS2</i> (N)	C/T	0.599	0.017	0.003	1.61×10^{-8}	All
rs9914578	17:1,951,886	<i>SMG6</i> (D,N); <i>N29617</i> (Q)	G/C	0.229	0.020	0.004	2.07×10^{-8}	All
rs977747	1:47,457,264	<i>TAL1</i> (N)	T/G	0.403	0.017	0.003	2.18×10^{-8}	All
rs9374842	6:120,227,364	<i>LOC285762</i> (N);	T/C	0.744	0.023	0.004	2.67×10^{-8}	EPB
rs4787491	16:29,922,838	<i>MAPK3</i> (D); <i>KCTD13</i> (D); <i>INO80E</i> (N); <i>TAOK2</i> (D); <i>YPEL3</i> (D); <i>DOC2A</i> (D); <i>FAM57B</i> (D)	G/A	0.510	0.022	0.004	2.70×10^{-8}	EPB
rs1441264	13:78,478,920	<i>MIR548A2</i> (N)	A/G	0.613	0.017	0.003	2.96×10^{-8}	All
rs17203016	2:207,963,763	<i>CREB1</i> (B,N); <i>KLF7</i> (B)	G/A	0.195	0.021	0.004	3.41×10^{-8}	All
rs16907751	8:81,538,012	<i>ZBTB10</i> (N)	C/T	0.913	0.047	0.009	3.89×10^{-8}	Men
rs13201877	6:137,717,234	<i>IFNGR1</i> (N); <i>OLIG3</i> (G)	G/A	0.140	0.024	0.004	4.29×10^{-8}	All
rs9540493	13:65,103,705	<i>MIR548X2</i> (N); <i>PCDH9</i> (D)	A/G	0.452	0.021	0.004	4.97×10^{-8}	EPB
rs1460676	2:164,275,935	<i>FIGN</i> (N)	C/T	0.179	0.021	0.004	4.98×10^{-8}	All
rs6465468	7:95,007,450	<i>ASB4</i> (B,N)	T/G	0.306	0.025	0.005	4.98×10^{-8}	Women
Previously identified loci								
rs6091540	20:50,521,269	<i>ZFP64</i> (N)	C/T	0.721	0.030	0.004	2.15×10^{-11}	Women
rs7715256	5:153,518,086	<i>GALNT10</i> (N)	G/T	0.422	0.017	0.003	8.85×10^{-9}	All
rs2176040	2:226,801,046	<i>LOC646736</i> (N); <i>IRS1</i> (B,Q)	A/G	0.365	0.024	0.004	9.99×10^{-9}	Men

SNP positions are reported according to Build 36 and their alleles are coded based on the positive strand. Alleles (effect/other), EAF, beta (β), s.e.m. and P values are based on the meta-analysis of GWAS I+II+Metabochip association data from the data set shown in the 'Analysis' column. EPB denotes European population-based studies, 'All' denotes all ancestries.

* Notable genes from biological relevance to obesity (B); copy number variation (C); DEPICT analyses (D); GRAIL results (G); BMI-associated variant is in strong LD ($r^2 \geq 0.7$) with a missense variant in the indicated gene (M); gene nearest to the index SNP (N); association and eQTL data converge to affect gene expression (Q).

Fig. 1b). Joint conditional analyses at two genomic regions separated by >500 kb (the *AGBL4-ELAVL4* regions on chr. 1, and the *ATP2A1-SBK1* regions on chr. 16), indicate that these pairs of signals may not be independent on extended LD.

Effects of BMI variants on other traits

We tested for associations between our 97 BMI-associated index SNPs and other metabolic phenotypes (Supplementary Tables 13–15 and Extended Data Figs 5 and 6). Thirteen of the twenty-three phenotypes

tested had significantly more SNPs with effects in the anticipated direction than expected by chance (Supplementary Table 16). These results corroborate the epidemiological relationships of BMI with metabolic traits. Whether this reflects a common genetic aetiology or a causal relationship of BMI on these traits requires further investigation.

Interestingly, some loci showed significant association with traits in the opposite direction than expected based on their phenotypic correlation with BMI (Extended Data Fig. 5). For example, at *HHIP*, the BMI-increasing allele is associated with decreased type 2 diabetes risk and

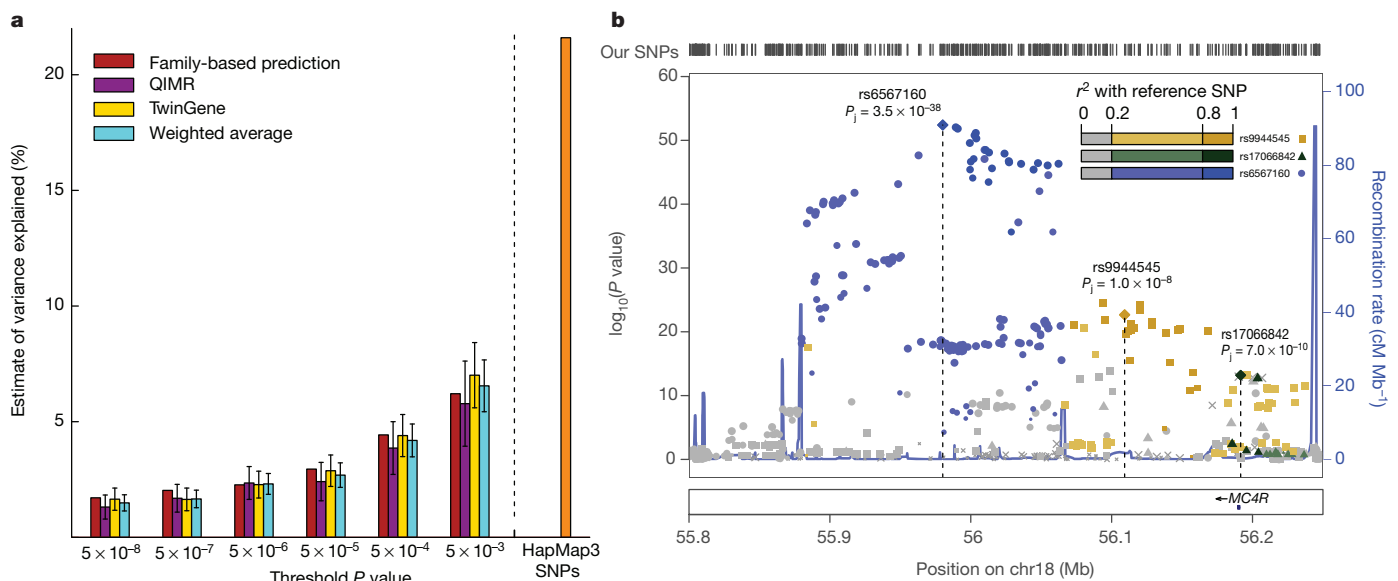


Figure 1 | Cumulative variance explained and example of secondary signals. **a**, The estimated variance in BMI explained by SNPs selected at a range of P values using unrelated individuals from the QIMR ($n = 3,924$; purple) and TwinGene ($n = 5,668$; gold), their weighted average (cyan), inferred from within-family prediction (red; Extended Data Fig. 2), and by all HapMap phase III SNPs in 16,275 unrelated individuals from the QIMR, TwinGene and ARIC

studies (orange). **b**, Plot of the region surrounding *MC4R* (ref. 36). SNP associations from the European sex-combined meta-analysis are plotted with joint conditional P values (P_j) indicated for the three conditionally significant signals. SNPs are shaded and shaped based on the index SNP with which they are in strongest LD (rs6567160 in blue, rs9944545 in yellow and rs17066842 in green).

Table 3 | Secondary signals reaching GWS by conditional analysis

SNP	Chr: position	Nearest gene	Alleles	EAF	β	s.e.m.	Variance explained	P value
rs1016287	2:59159129	<i>LINC01122</i>	T/C	0.294	0.023	0.003	0.021%	2.62×10^{-11}
rs4671328	2:58788786	<i>LINC01122</i>	T/G	0.457	0.021	0.004	0.021%	2.73×10^{-8}
rs758747	16:3567359	<i>NLRC3</i>	T/C	0.241	0.022	0.004	0.018%	2.00×10^{-9}
rs879620	16:3955730	<i>ADCY9</i>	T/C	0.620	0.024	0.004	0.027%	2.17×10^{-9}
rs12446632	16:19842890	<i>GPRC5B</i>	G/A	0.860	0.036	0.005	0.031%	1.06×10^{-14}
rs11074446	16:20162624	<i>GP2</i>	T/C	0.867	0.029	0.005	0.019%	1.71×10^{-10}
rs6567160	18:55980115	<i>MC4R</i>	C/T	0.233	0.048	0.004	0.084%	3.52×10^{-38}
rs17066842	18:56191604	<i>MC4R</i>	G/A	0.960	0.051	0.008	0.020%	6.99×10^{-10}
rs9944545	18:56109224	<i>MC4R</i>	T/C	0.296	0.020	0.004	0.017%	1.01×10^{-8}
rs11030104	11:27641093	<i>BDNF</i>	A/G	0.791	0.051	0.004	0.087%	1.26×10^{-34}
rs10835210	11:27652486	<i>BDNF</i>	C/A	0.570	0.020	0.004	0.020%	1.25×10^{-8}

SNP positions are reported according to Build 36 and their alleles are coded based on the positive strand. Alleles (effect/other), EAF, estimated beta (β), s.e.m., explained variance, and P values from GCTA. First row at each locus represents lead signal, other row(s) represent secondary signals.

higher high-density lipoprotein cholesterol (HDL). At *LOC646736* and *IRSI*, the BMI-increasing allele is associated with reduced risk of coronary artery disease (CAD) and diabetic nephropathy, decreased triglyceride levels, increased HDL, higher adiponectin, and lower fasting insulin. This may be due to increased subcutaneous fat and possible production of metabolic mediators protective against the development of metabolic disease despite increased adiposity⁸. These unexpected associations may help us to understand better the complex pathophysiology underlying these traits, and may indicate benefits or side effects if these regions contain targets of therapeutic intervention. Furthermore, of our 97 GWS loci, 35 (binomial $P = 0.0019$) were in high LD ($r^2 > 0.7$) with one or more GWS SNPs in the National Human Genome Research Institute (NHGRI) GWAS catalogue ($P < 5 \times 10^{-8}$), even after removing anthropometric trait-associated SNPs. These SNPs were associated not only with cardiometabolic traits, but also with schizophrenia, smoking behaviour, irritable bowel syndrome, and Alzheimer's disease (Supplementary Table 17a, b).

BMI tissues, biological pathways and gene sets

We anticipated the expanded sample size would not only identify additional BMI-associated variants, but also more clearly highlight the biology implicated by genetic studies of BMI. By applying multiple complementary methods, we identified biologically relevant tissues, pathways and gene sets, and highlighted potentially causal genes at associated loci. These approaches included systematic methods incorporating diverse data types, including the novel approach, Data-driven Expression Prioritized Integration for Complex Traits (DEPICT)²¹, and extensive manual review of the literature.

DEPICT used 37,427 human gene expression microarray samples to identify tissues and cell types in which genes near BMI-associated SNPs are highly expressed, and then tested for enrichment of specific tissues by comparing results with randomly selected loci matched for gene density. In total, 27 out of 31 significantly enriched tissues were in the central nervous system (CNS) (out of 209 tested; Fig. 2a and Supplementary Table 18). Current results are not sufficient to isolate specific brain regions important in regulating BMI. However, we observe enrichment not only in the hypothalamus and pituitary gland—key sites of central appetite regulation—but even more strongly in the hippocampus and limbic system, tissues that have a role in learning, cognition, emotion and memory.

As a complementary approach, we examined overlap of associated variants at the 97 loci ($r^2 > 0.7$ with the lead SNP) with five regulatory marks found in most of the 14 selected cell types from brain, blood, liver, pancreatic islet and adipose tissue from the ENCODE Consortium²² and Roadmap Epigenomics Project²³ (Supplementary Table 19a–c). We found evidence of enrichment ($P < 1.2 \times 10^{-3}$) in 24 out of 41 data sets examined. The strongest enrichment was observed with promoter (histone 3 Lys 4 trimethylation (H3K4me3), histone 3 Lys 9 acetylation (H3K9ac)) and enhancer (H3K4me1, H3K27ac) marks detected in mid-frontal lobe, anterior caudate, astrocytes and substantia nigra, supporting neuronal tissues in BMI regulation.

To identify pathways or gene sets implicated by the BMI-associated loci, we first used Meta-Analysis Gene-set Enrichment of variant Associations (MAGENTA)²⁴, which takes as input pre-annotated gene sets, and then tests for overrepresentation of gene set genes at BMI-associated loci. We found enrichment (false discovery rate (FDR) < 0.05) of seven gene sets, including neurotrophin signalling. Other highlighted gene sets related to general growth and patterning: basal cell carcinoma, acute myeloid leukaemia, and hedgehog signalling (Supplementary Table 20a, b).

Second, we used DEPICT, that uses predefined gene sets reconstituted using coexpression data, to perform gene set enrichment analysis. After merging highly correlated gene sets, nearly 500 gene sets were significantly enriched (FDR < 0.05) for genes in BMI-associated loci (Fig. 2b and Supplementary Table 21a, b). The most strongly enriched gene sets highlight potentially novel pathways in the CNS. These include gene sets related to synaptic function, long-term potentiation and neurotransmitter signalling (glutamate signalling in particular, but also noradrenaline, dopamine and serotonin release cycles, and GABA (γ -aminobutyric acid) receptor activity; Fig. 2c). Potentially relevant mouse behavioural phenotypes, such as physical activity and impaired coordination were also highly enriched (Fig. 2b and Supplementary Table 21a). Several gene sets previously linked to obesity, such as integration of energy metabolism, polyphagia, secretion and action of insulin and related hormones (for example, 'regulation of insulin secretion by glucagon-like peptide 1' and 'glucagon signalling in metabolic regulation'), mTOR signalling (which affects cell growth in response to nutrient intake via insulin and growth factors²⁵), and gene sets overlapping the neurotrophin signalling pathway identified by MAGENTA were also enriched, although not as significantly as other CNS processes (Fig. 2d). DEPICT also identified significant enrichment for additional cellular components and processes: calcium channels, MAP kinase activity, chromatin organization and modification, and ubiquitin ligases.

Third, we manually reviewed literature related to all 405 genes within 500 kb and $r^2 > 0.2$ of the 97 index SNPs. We classified these genes into one or more biological categories, and observed 25 categories containing three or more genes (Supplementary Table 22). The largest category comprised genes involved in neuronal processes, including monogenic obesity genes involved in hypothalamic function and energy homeostasis, and genes involved in neuronal transmission and development. Other processes highlighted by the manual literature review included glucose and lipid homeostasis and limb development, which were less notable in the above methods, but may still be related to the underlying biology of BMI.

To identify specific genes that may account for BMI association, we considered each of the following to represent supportive evidence for a gene within a locus: (1) the gene nearest the index SNP²⁶; (2) genes containing missense, nonsense or copy number variants, or a *cis*-expression quantitative trait locus (eQTL) in LD with the index SNP; (3) genes prioritized by integrative methods implemented in DEPICT; (4) genes prioritized by connections in published abstracts by GRAIL (Gene Relationships Across Implicated Loci)²⁷; or (5) genes biologically related

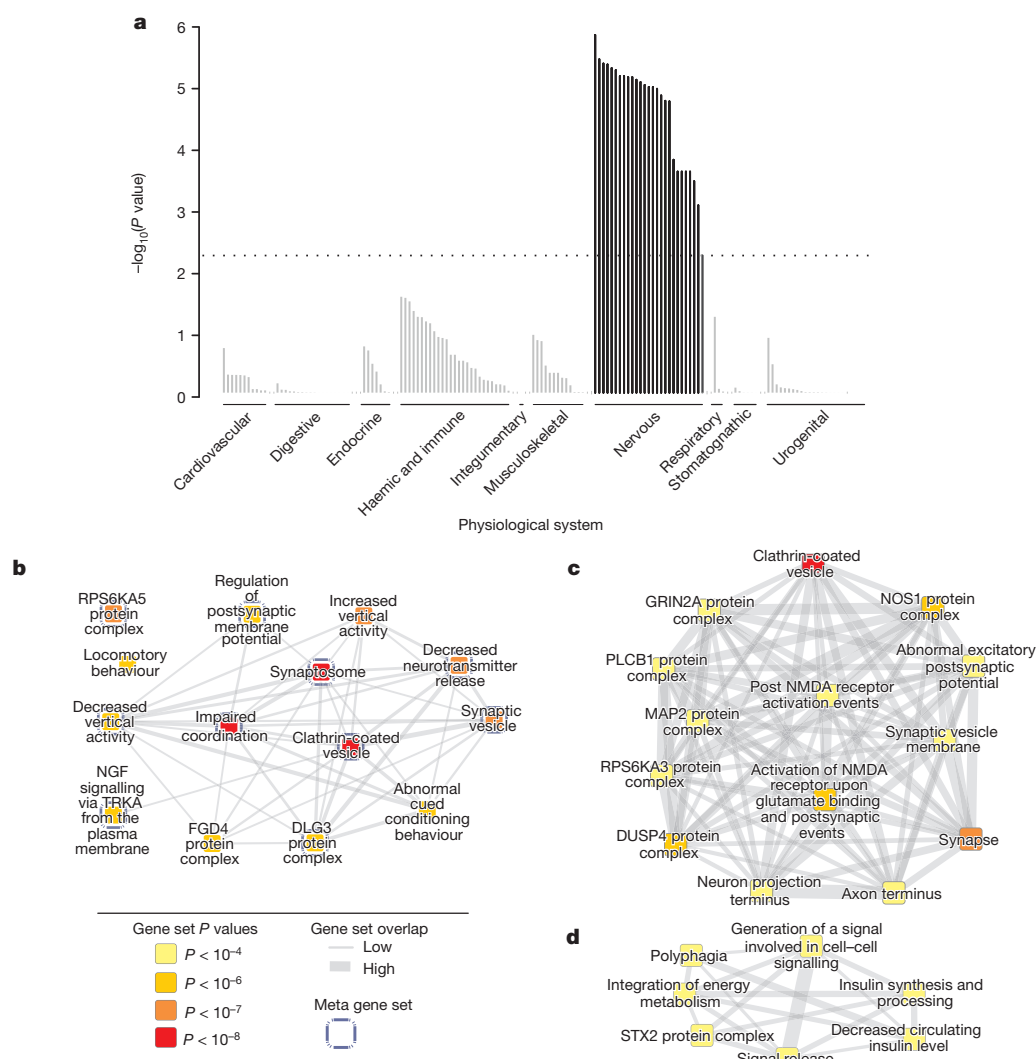


Figure 2 | Tissues and reconstituted gene sets significantly enriched for genes within BMI-associated loci.

a, DEPICT predicts genes within BMI-associated loci ($P < 5 \times 10^{-4}$) are enriched for expression in the brain and central nervous system. Tissues are sorted by physiological system and significantly enriched tissues are in black; the dotted line represents statistically significant enrichment. **b**, The gene sets most significantly enriched for BMI-associated loci by DEPICT ($P < 10^{-6}$, $FDR < 4 \times 10^{-4}$). Nodes represent reconstituted gene sets and are colour-coded by P value. Edge thickness between nodes is proportional to degree of gene overlap as measured by the Jaccard index. Nodes with gene overlap greater than 25% were collapsed into a single 'meta-node' (blue border). **c**, The nodes contained within the most enriched meta-node, 'clathrin-coated vesicle', which shares genes with other gene sets relevant to glutamate signalling and synapse biology. **d**, The 'generation of a signal involved in cell-cell signalling' meta-node represents several overlapping gene sets relevant to obesity and energy metabolism (gene sets with $P < 4 \times 10^{-3}$, $FDR < 0.05$ shown). For the complete list of enriched gene sets refer to Supplementary Table 21a.

to obesity, related metabolic disease, or energy expenditure based on manual literature review (Tables 1 and 2, Extended Data Tables 2–4 and Supplementary Tables 23–25). We first focused on the 64 genes in associated loci with more than one consistent line of supporting evidence. As expected, many of these genes overlap with CNS processes, including synaptic function, cell–cell adhesion, and glutamate signalling (*ELAVL4*, *GRID1*, *CADM2*, *NRXN3*, *NEGR1* and *SCG3*), cause monogenic obesity syndromes (*MC4R*, *BDNF*, *BBS4* and *POMC*), or function in extreme/early onset obesity in humans and mouse models (*SH2B1* and *NEGR1*)^{6,28,29}. Other genes with several lines of supporting evidence are related to insulin secretion and action, energy metabolism, lipid biology, and/or adipogenesis (*TCF7L2*, *GIPR*, *IRS1*, *FOXO3*, *ASB4*, *RPTOR*, *NPC1*, *CREB1*, *FAM57B*, *APOBR* and *HSD17B12*), encode RNA binding/processing proteins (*PTBP2*, *ELAVL4*, *CELFI* and possibly *RALYL*), are in the MAP kinase signalling pathway (*MAP2K5* and *MAPK3*), or regulate cell proliferation or cell survival (*FAIM2*, *PARK2* and *OLFM4*). Although we cannot be certain that any individual gene is related to the association at a given locus, the strong enrichment of pathways among genes within associated loci argues for a causal role for these pathways, prioritizes specific genes for follow-up experiments, and provides the strongest genetic evidence so far for a role of particular biological and CNS processes in the regulation of human body mass.

Discussion

Our meta-analysis of nearly 340,000 individuals identified 97 GWS loci associated with BMI, 56 of which are novel. These loci account for 2.7% of the variation in BMI, and suggest that as much as 21% of BMI

variation can be accounted for by common genetic variation. Our analyses provide robust evidence to implicate particular genes and pathways affecting BMI, including synaptic plasticity and glutamate receptor activity—pathways that respond to changes in feeding and fasting, are regulated by key obesity-related molecules such as BDNF and MC4R, and impinge on key hypothalamic circuits^{30–32}. These pathways also overlap with one of the several proposed mechanisms of action of topiramate, a component of one of two weight-loss drugs approved by the US Food and Drug Administration^{33,34}. This observation suggests that the relevant site of action for this drug may be glutamate receptor activity, supporting the idea that these genes and pathways could reveal more targets for weight-loss therapies. BMI-associated loci also overlap with genes and pathways implicated in neurodevelopment (Supplementary Tables 21 and 22). Finally, consistent with previous work and findings from monogenic obesity syndromes, we confirm a role for the CNS—particularly genes expressed in the hypothalamus—in the regulation of body mass.

Examining the genes at BMI-associated loci in the context of gene expression, molecular pathways, eQTL results, mutational evidence and genomic location provides several complementary avenues through which to prioritize genes for relevance in BMI biology. Genes such as *NPC1* and *ELAVL4* are implicated by many lines of evidence (literature, mutational, eQTL and DEPICT) and become strong candidate genes in their respective locations. It is important to recognize that pathway methods and literature reviews are limited by current data sets and knowledge, and thus provide only a working model of obesity biology. For example, little is known about host genetic factors that regulate the

microbiome. Variation in immune-related genes such as *TLR4* could presumably exert an influence on obesity through the microbiome³⁵. Together, our results underscore the heterogeneous aetiology of obesity and its links with several related metabolic diseases and processes.

BMI variants are generally associated with related cardiometabolic traits in accord with established epidemiological relationships. This could be due to shared genetic effects or to other causes of cross-phenotypic correlations. However, some BMI-associated variants have effects on related traits counter to epidemiological expectations. Once better understood, these mechanisms may not only help to explain why not all obese individuals develop related metabolic diseases, but also suggest possible mechanisms to prevent development of metabolic disease in those who are already obese.

Larger studies of common genetic variation, studies of rare variation (including those based on imputation, exome chips and sequencing), and improved computational tools will continue to identify genetic variants associated with BMI and help to further refine the biology of obesity. The 97 loci identified here represent an important step in understanding the physiological mechanisms leading to obesity. These findings strengthen the connection between obesity and other metabolic diseases, enhance our appreciation of the tissues, physiological processes, and molecular pathways that contribute to obesity, and will guide future research aimed at unravelling the complex biology of obesity.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 20 November 2013; accepted 23 December 2014.

- Maes, H. H., Neale, M. C. & Eaves, L. J. Genetic and environmental factors in relative body weight and human adiposity. *Behav. Genet.* **27**, 325–351 (1997).
- Visscher, P. M., Brown, M. A., McCarthy, M. I. & Yang, J. Five years of GWAS discovery. *Am. J. Hum. Genet.* **90**, 7–24 (2012).
- Zaitlen, N. *et al.* Using extended genealogy to estimate components of heritability for 23 quantitative and dichotomous traits. *PLoS Genet.* **9**, e1003520 (2013).
- Fall, T. & Ingelsson, E. Genome-wide association studies of obesity and metabolic syndrome. *Mol. Cell. Endocrinol.* **382**, 740–757 (2014).
- Speliotes, E. K. *et al.* Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nature Genet.* **42**, 937–948 (2010).
- Willer, C. J. *et al.* Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. *Nature Genet.* **41**, 25–34 (2009).
- Voight, B. F. *et al.* The metabolochip, a custom genotyping array for genetic studies of metabolic, cardiovascular, and anthropometric traits. *PLoS Genet.* **8**, e1002793 (2012).
- Kilpeläinen, T. O. *et al.* Genetic variation near *IRS1* associates with reduced adiposity and an impaired metabolic profile. *Nature Genet.* **43**, 753–760 (2011).
- Bradfield, J. P. *et al.* A genome-wide association meta-analysis identifies new childhood obesity loci. *Nature Genet.* **44**, 526–531 (2012).
- Monda, K. L. *et al.* A meta-analysis identifies new loci associated with body mass index in individuals of African ancestry. *Nature Genet.* **45**, 690–696 (2013).
- Berndt, S. I. *et al.* Genome-wide meta-analysis identifies 11 new loci for anthropometric traits and provides insights into genetic architecture. *Nature Genet.* **45**, 501–512 (2013).
- Guo, Y. *et al.* Gene-centric meta-analyses of 108 912 individuals confirm known body mass index loci and reveal three novel signals. *Hum. Mol. Genet.* **22**, 184–201 (2013).
- Wood, A. R. *et al.* Defining the role of common variation in the genomic and biological architecture of adult human height. *Nature Genet.* **46**, 1173–1186 (2014).
- Maller, J. B. *et al.* Bayesian refinement of association signals for 14 loci in 3 common diseases. *Nature Genet.* **44**, 1294–1301 (2012).
- Wakefield, J. A Bayesian measure of the probability of false discovery in genetic epidemiology studies. *Am. J. Hum. Genet.* **81**, 208–227 (2007).
- Peters, U. *et al.* A systematic mapping approach of 16q12.2/FTO and BMI in more than 20,000 African Americans narrows in on the underlying functional variation: results from the Population Architecture using Genomics and Epidemiology (PAGE) study. *PLoS Genet.* **9**, e1003171 (2013).
- Juster, F. T. & Suzman, R. An overview of the Health and Retirement Study. *J. Hum. Resour.* **30**, S7–S56 (1995).
- Bouchonville, M. *et al.* Weight loss, exercise or both and cardiometabolic risk factors in obese older adults: results of a randomized controlled trial. *Int. J. Obes.* **38**, 423–431 (2013).
- Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).
- Yang, J. *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nature Genet.* **44**, 369–375 (2012).
- Pers, T. *et al.* Biological interpretation of genome-wide association studies using predicted gene functions. *Nat. Commun.* **5**, 5890 (2014).
- The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
- Bernstein, B. E. *et al.* The NIH Roadmap Epigenomics Mapping Consortium. *Nature Biotechnol.* **28**, 1045–1048 (2010).
- Segrè, A. V., Groop, L., Mootha, V. K., Daly, M. J. & Altshuler, D. Common inherited variation in mitochondrial genes is not enriched for associations with type 2 diabetes or related glycemic traits. *PLoS Genet.* **6**, e1001058 (2010).
- Wullschlegel, S., Loewith, R. & Hall, M. N. TOR signaling in growth and metabolism. *Cell* **124**, 471–484 (2006).
- Lango Allen, H. *et al.* Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* **467**, 832–838 (2010).
- Raychaudhuri, S. *et al.* Identifying relationships among genomic disease regions: predicting genes at pathogenic SNP associations and rare deletions. *PLoS Genet.* **5**, e1000534 (2009).
- Mägi, R. *et al.* Contribution of 32 GWAS-identified common variants to severe obesity in European adults referred for bariatric surgery. *PLoS ONE* **8**, e70735 (2013).
- Lee, A. W. *et al.* Functional inactivation of the genome-wide association study obesity gene neuronal growth regulator 1 in mice causes a body mass phenotype. *PLoS ONE* **7**, e11537 (2012).
- Yang, Y., Atasoy, D., Su, H. H. & Sternson, S. M. Hunger states switch a flip-flop memory circuit via a synaptic AMPK-dependent positive feedback loop. *Cell* **146**, 992–1003 (2011).
- Wu, Q., Clark, M. S. & Palmiter, R. D. Deciphering a neuronal circuit that mediates appetite. *Nature* **483**, 594–597 (2012).
- Shen, Y., Fu, W. Y., Cheng, E. Y., Fu, A. K. & Ip, N. Y. Melanocortin-4 receptor regulates hippocampal synaptic plasticity through a protein kinase A-dependent mechanism. *J. Neurosci.* **33**, 464–472 (2013).
- Gibbs, J. W., III, Sombati, S., DeLorenzo, R. J. & Coulter, D. A. Cellular actions of topiramate: blockade of kainate-evoked inward currents in cultured hippocampal neurons. *Epilepsia* **41** (suppl. 1), S10–S16 (2000).
- Poulsen, C. F. *et al.* Modulation by topiramate of AMPA and kainate mediated calcium influx in cultured cerebral cortical, hippocampal and cerebellar neurons. *Neurochem. Res.* **29**, 275–282 (2004).
- Henaoui-Meija, J. *et al.* Inflammation-mediated dysbiosis regulates progression of NAFLD and obesity. *Nature* **482**, 179–185 (2012).
- Pruim, R. J. *et al.* LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* **26**, 2336–2337 (2010).

Supplementary Information is available in the online version of the paper.

Acknowledgements A full list of acknowledgements can be found in the Supplementary Information.

Author Contributions A full list of author contributions can be found in the Supplementary Information.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to E.K.S. (espeliot@med.umich.edu), R.J.F.L. (ruth.loos@mssm.edu), and J.N.H. (joel@broadinstitute.org).

Adam E. Locke^{1*}, Bratati Kahali^{2*}, Sonja I. Berndt^{3*}, Anne E. Justice^{4*}, Tune H. Pers^{5,6,7,8*}, Felix R. Day⁹, Corey Powell², Sailaja Vedantam^{5,6}, Martin L. Buchkovich¹⁰, Jian Yang^{11,12}, Damien C. Croteau-Chonka^{10,13}, Tonu Esko^{5,6,7,14}, Tove Fall^{15,16,17}, Teresa Ferreira¹⁸, Stefan Gustafsson^{16,17}, Zoltán Kutalik^{19,20,21}, Jian'an Luan⁹, Reedik Mägi^{14,18}, Joshua C. Randall^{18,22}, Thomas W. Winkler²³, Andrew R. Wood²⁴, Tsegaselassie Workalemahu²⁵, Jessica D. Faul²⁶, Jennifer A. Smith²⁷, Jing Hua Zhao⁹, Wei Zhao²⁷, Jin Chen²⁸, Rudolf Fehrmann²⁹, Åsa K. Hedman^{16,17,18}, Juha Karjalainen²⁹, Ellen M. Schmidt³⁰, Devin Absher³¹, Najaf Amin³², Denise Anderson³³, Marian Beekman^{34,35}, Jennifer L. Bolton³⁶, Jennifer L. Bragg-Gresham^{1,37}, Steven Buyske^{38,39}, Ayse Demirkan^{32,40}, Guohong Deng^{41,42,43}, Georg B. Ehret^{44,45}, Bjarke Feenstra⁴⁶, Mary F. Feitosa⁴⁷, Krista Fischer¹⁴, Anuj Goel^{18,48}, Jian Gong⁴⁹, Anne U. Jackson⁵¹, Stavroula Kanoni⁵⁰, Marcus E. Kleber^{51,52}, Kati Kristiansson⁵³, Unhee Lim⁵⁴, Vaneet Lotay⁵⁵, Massimo Mangino⁵⁶, Irene Mateo Leach⁵⁷, Carolina Medina-Gomez^{58,59,60}, Sarah E. Medland⁶¹, Michael A. Nalls⁶², Cameron D. Palmer^{5,6}, Dorota Pasko²⁴, Sonali Pechlivanis⁵³, Marjolijn J. Peters^{58,60}, Inga Prokopenko^{18,64,65}, Dmitry Shungin^{66,67,68}, Alena Stancáková⁶⁹, Rona J. Strawbridge⁷⁰, Yun Ju Sung⁷¹, Toshiko Tanaka⁷², Alexander Teumer⁷³, Stella Trompet^{74,75}, Sander W. van der Laan⁷⁶, Jessica van Setten⁷⁷, Jana V. Van Vliet-Ostapchouk⁷⁸, Zhaoiming Wang^{3,79}, Loic Yengo^{80,81,82}, Weihua Zhang^{41,83}, Aaron Isaacs^{32,84}, Eva Albrecht⁸⁵, Johan Ärnlöv^{16,17,86}, Gillian M. Arscott⁸⁷, Antony P. Attwood^{88,89}, Stefania Bandinelli⁹⁰, Amy Barrett⁶⁴, Isabelita N. Bas⁹¹, Claire Bellis^{92,93}, Amanda J. Bennett⁶⁴, Christian Berne⁹⁴, Roza Blagieva⁹⁵, Matthias Blüher^{96,97}, Stefan Böhringer^{34,98}, Lori L. Bonnycastle⁹⁹, Yvonne Böttcher⁹⁶, Heather A. Boyd⁴⁶, Marcel Bruinenberg¹⁰⁰, Ida H. Caspersen¹⁰¹, Yli-Der Ida Chen^{102,103}, Robert Clarke¹⁰⁴, E. Warwick Daw⁴⁷, Anton J. M. de Craen⁷⁵, Graciela Delgado⁵¹, Maria Dimitriou¹⁰⁵, Alex S. F. Doney¹⁰⁶, Niina Eklund^{53,107}, Karol Estrada^{6,60,108}, Elodie Eury^{80,81,82}, Lasse Folkersen⁷⁰, Ross M. Fraser³⁶, Melissa E. Garcia¹⁰⁹, Frank Geller⁴⁶, Vilmantas Giedraitis¹¹⁰, Bruna Gigante¹¹¹, Alan S. Go¹¹², Alain Golay¹¹³, Alison H. Goodall^{114,115}, Scott D. Gordon⁶¹, Mathias Gorski^{23,116}, Hans-Jörgen Grabe^{117,118}, Harald Grallert^{85,119,120}, Tanja B. Grammer⁵¹, Jürgen Gräßler¹²¹, Henrik Grönberg¹⁵, Christopher J. Groves⁶⁴, Gaëlle Gusto¹²², Jeffrey

- Haessler⁴⁹, Per Hall¹⁵, Toomas Haller¹⁴, Goran Hallmans¹²³, Catharina A. Hartman¹²⁴, Maija Hassinen¹²⁵, Caroline Hayward¹²⁶, Nancy L. Heard-Costa^{127,128}, Quinta Helmer^{34,98,129}, Christian Hengstenberg^{130,131}, Oddgeir Holmen¹³², Jouke-Jan Hottenga¹³³, Alan L. James^{134,135}, Janina M. Jeff⁵⁵, Åsa Johansson¹³⁶, Jennifer Jolley^{88,89}, Thorhildur Juliusdottir¹⁸, Leena Kinnunen⁵³, Wolfgang Koenig⁵², Markku Koskenvuo¹³⁷, Wolfgang Kratzer¹³⁸, Jaana Laitinen¹³⁹, Claudia Lamina¹⁴⁰, Karin Leander¹¹¹, Nanette R. Lee⁹¹, Peter Lichtner¹⁴¹, Lars Lind¹⁴², Jaana Lindström⁵³, Ken Sin Lo¹⁴³, Stéphane Lobbes^{80,81,82}, Roberto Lorbeer¹⁴⁴, Yingchang Lu^{55,145}, François Mach⁴⁵, Patrik K. E. Magnusson¹⁵, Anubha Mahajan¹⁸, Wendy L. McArdle¹⁴⁶, Stela McLachlan³⁶, Cristina Menni⁵⁶, Sigrun Merger⁹⁵, Evelin Mihailov¹⁴⁷, Lili Milani¹⁴, Alireza Moayyeri^{56,148}, Keri L. Monda^{4,149}, Mario A. Morken⁹⁹, Antonella Mulas¹⁵⁰, Gabriele Müller¹⁵¹, Martina Müller-Nurasyid^{85,130,152,153}, Arthur W. Musk¹⁵⁴, Ramaiah Nagaraja¹⁵⁵, Markus M. Nöthen^{56,157}, Ilja M. Nolte¹⁵⁸, Stefan Pilz^{159,160}, Nigel W. Rayner^{18,22,64}, Frida Renstrom⁶⁶, Rainer Rettig¹⁶¹, Janina S. Ried⁸⁵, Stephan Ripke^{108,162}, Neil R. Robertson^{18,64}, Lynda M. Rose¹⁶³, Serena Sanna¹⁵⁰, Hubert Scharnagl¹⁶⁴, Salome Scholtens¹⁰⁰, Fredrick R. Schumacher¹⁶⁵, William R. Scott^{41,83}, Thomas Seufferlein¹³⁸, Jianxin Shi¹⁶⁶, Albert Vernon Smith^{167,168}, Joanna Smolonska^{29,169}, Alice V. Stanton¹⁷⁰, Valgerdur Steinthorsdottir¹⁷¹, Kathleen Stirrups^{22,50}, Heather M. Stringham¹, Johan Sundström¹⁴², Morris A. Swertz²⁹, Amy J. Swift⁹⁹, Ann-Christine Syvänen^{16,172}, Sian-Tsung Tan^{41,173}, Bamidele O. Tayo¹⁷⁴, Barbara Thorand^{120,175}, Gudmar Thorleifsson¹⁷¹, Jonathan P. Tyrer¹⁷⁶, Hae-Won Uh^{34,98}, Liesbeth Vandenput¹⁷⁷, Frank C. Verhulst¹⁷⁸, Sita H. Vermeulen^{179,180}, Niek Verweij⁵⁷, Judith M. Vonk¹⁶⁹, Lindsay L. Waite³¹, Helen R. Warren¹⁸¹, Dawn Waterworth¹⁸², Michael N. Weedon²⁴, Lynne R. Wilkens⁵⁴, Christina Willenborg^{83,184}, Tom Wilsaard¹⁸⁵, Mary K. Wojczynski⁴⁷, Andrew Wong¹⁸⁶, Alan F. Wright¹²⁶, Qunyan Zhang⁴⁷, The LifeLines Cohort Study⁴⁷, Eoin P. Brennan¹⁸⁷, Murim Choi¹⁸⁸, Zari Dastani¹⁸⁹, Alexander W. Drong¹⁸, Per Eriksson⁷⁰, Anders Franco-Cereceda¹⁹⁰, Jesper R. Gadin⁷⁰, Ali G. Gharavi¹⁹¹, Michael E. Goddard^{192,193}, Robert E. Handsaker⁶⁷, Jinyan Huang^{194,195}, Fredrik Karpe^{64,196}, Sekar Kathiresan^{6,197}, Sarah Keildson¹⁸, Krzysztof Kiryluk¹⁹¹, Michiaki Kubo¹⁹⁸, Jong-Young Lee¹⁹⁹, Liming Liang^{194,200}, Richard P. Lifton²⁰¹, Baoshan Ma^{194,202}, Steven A. McCarrall^{67,162}, Amy J. McKnight²⁰³, Josine L. Min¹⁴⁶, Miriam F. Moffatt¹⁷³, Grant W. Montgomery⁶¹, Joanne M. Murabito^{127,204}, George Nicholson^{205,206}, Dale R. Nyholt^{61,207}, Yukunori Okada^{208,209}, John R. B. Perry^{182,246}, Rajkumar Dorajoo²¹⁰, Eva Reinmaa¹⁴, Rany M. Salem^{5,6,7}, Niina Sandholm^{211,212,213}, Robert A. Scott⁹, Lisette Stolk^{34,60}, Atsushi Takahashi²⁰⁸, Toshihiro Tanaka^{209,214,215}, Ferdinand M. van 't Hooft⁷⁰, Anna A. E. Vinkhuysen¹¹, Harm-Jan Westra²⁹, Wei Zheng²¹⁶, Krina T. Zondervan^{18,217}, The ADIPOGen Consortium[†], The AGEN-BMI Working Group[†], The CARDIOGRAMplusC4D Consortium[†], The CKDGen Consortium[†], The GLGC[†], The ICBP[†], The MAGIC Investigators[†], The MuTHER Consortium[†], The MIGN Consortium[†], The PAGE Consortium[†], The ReproGen Consortium[†], The GENIE Consortium[†], The International Endogene Consortium[†], Andrew C. Heath²¹⁸, Dominique Arveiler²¹⁹, Stephan J. L. Bakker²²⁰, John Beilby^{87,221}, Richard N. Bergman²²², John Blangero⁹², Pascal Bovet^{223,224}, Harry Campbell³⁶, Mark J. Caulfield¹⁸¹, Giancarlo Cesana²²⁵, Aravinda Chakravarti⁴⁴, Daniel I. Chasman^{163,226}, Peter S. Chines⁹⁹, Francis S. Collins⁹⁹, Dana C. Crawford^{227,228}, L. Adrienne Cupples^{127,229}, Daniele Cusi^{230,231}, John Danesh³, Ulf de Faire¹¹¹, Hester M. den Ruijter^{76,233}, Anna F. Dominiczak²³⁴, Raimund Erbel²³⁵, Jeanette Erdmann^{183,184}, Johan G. Eriksson^{53,236,237}, Martin Farrall^{18,48}, Stephan B. Felix^{238,239}, Ele Ferrannini^{240,241}, Jean Ferrières²⁴², Ian Ford²⁴³, Nita G. Forouhi⁹, Terrence Forrester²⁴⁴, Oscar H. Franco^{58,59}, Ron T. Gansevoort²²⁰, Pablo V. Gejman²⁴⁵, Christian Gieger⁸⁵, Omri Gottesman⁵⁵, Vilhelm Gudnason^{167,168}, Ulf Gyllenstein¹³⁶, Alistair S. Hall²⁴⁶, Tamara B. Harris¹⁰⁹, Andrew T. Hattersley²⁴⁷, Andrew A. Hicks²⁴⁸, Lucia A. Hindorf²⁴⁹, Aaron D. Hingorani²⁵⁰, Albert Hofman^{58,59}, Georg Homuth⁷³, G. Kees Hovingh²⁵¹, Steve E. Humphries²⁵², Steven C. Hunt²⁵³, Elina Hyppönen^{254,255,256,257}, Thomas Illig^{19,258}, Kevin B. Jacobs^{3,79}, Marjo-Riitta Jarvelin^{83,259,260,261,263,263}, Karl-Heinz Jöckel⁶³, Berit Johansen¹⁰¹, Pekka Jousilahti⁵³, J. Wouter Jukema^{74,264,265}, Antti M. Jula⁵³, Jaakko Kaprio^{53,107,137}, John J. Kastelein²⁵¹, Sirkka M. Keinänen-Kiukkaanniemi^{266,266}, Lambertus A. Kiemeny^{179,267}, Paul Knekt⁵³, Jaspal S. Kooner^{41,173,268}, Charles Kooperberg⁴⁹, Peter Kovacs^{96,97}, Aldi T. Kraja⁴⁷, Meena Kumari^{269,270}, Johanna Kuusisto²⁷¹, Timo A. Lakka^{125,272,273}, Claudia Langenberg^{9,269}, Loïc Le Marchand⁵⁴, Terho Lehtimäki²⁷⁴, Valeriya Lyssenko^{275,276}, Satu Männistö⁵³, André Marette^{277,278}, Tara C. Matise³⁹, Colin A. McKenzie²⁴⁴, Barbara McKnight²⁷⁹, Frans L. Moll²⁸⁰, Andrew D. Morris¹⁰⁶, Andrew P. Morris^{14,18,281}, Jeffrey C. Murray²⁸², Mari Nelis¹⁴, Claes Ohlsson¹⁷⁷, Albertine J. Oldehinkel¹²⁴, Ken K. Ong^{9,186}, Pamela A. F. Madden²¹⁸, Gerard Pasterkamp⁷⁶, John F. Peden²⁸³, Annette Peters^{119,130,175}, Dirkje S. Postma²⁸⁴, Peter P. Pramstaller^{248,285}, Jackie F. Price³⁶, Lu Qi^{3,25}, Olli T. Raitakari^{286,287}, Tuomo Rankinen²⁸⁸, D. C. Rao^{47,71,218}, Treva K. Rice^{71,218}, Paul M. Ridker^{163,228}, John D. Rioux^{143,289}, Marylyn D. Ritchie²⁹⁰, Igor Rudan^{36,291}, Veikko Salomaa⁵³, Nilesh J. Samani^{114,115}, Jouko Samraies²⁹², Mark A. Sarzynski²⁸⁸, Heribert Schunkert^{30,131}, Peter E. H. Schwarz^{121,293}, Peter Severe²⁹⁴, Alan R. Shuldiner^{295,296,297}, Juha Sinisalo²⁹⁸, Ronald P. Stolk¹⁶⁹, Konstantin Strauch^{85,153}, Anke Tönjes^{96,97}, David-Alexandre Tréguet^{299,300,301}, Angelo Tremblay³⁰², Elena Tremoli³⁰³, Jarmo Virtamo⁵³, Marie-Claude Vohl^{278,304}, Uwe Völker^{73,239}, Gérard Waeber³⁰⁵, Gonneke Willemssen¹³³, Jacqueline C. Witteman⁵⁹, M. Carola Zillikens^{58,60}, Linda S. Adair³⁰⁶, Philippe Amouyel³⁰⁷, Folkert W. Asselbergs^{250,264,308}, Themistocles L. Assimes³⁰⁹, Murielle Bochud^{223,224}, Bernhard O. Boehm^{310,311}, Eric Boerwinkle³¹², Stefan R. Bornstein¹²¹, Erwin P. Bottinger⁵⁵, Claude Bouchard²⁸⁸, Stéphane Cauchi^{80,81,82}, John C. Chambers^{41,83,268}, Stephen J. Chanock¹⁸³, Richard S. Cooper¹⁷⁴, Paul I. W. de Bakker^{77,313,314}, George Dedoussis¹⁰⁵, Luigi Ferrucci⁷², Paul W. Franks¹⁵, Philippe Froguel^{65,80,81,82}, Leif C. Groop^{107,276}, Christopher A. Haiman¹⁶⁵, Anders Hamsten⁷⁰, Jennie Hui^{87,221,315}, David J. Hunter^{13,25,194}, Kristian Hveem¹³², Robert C. Kaplan³¹⁶, Mika Kivimäki²⁶⁹, Diana Kuh¹⁸⁶, Markku Laakso²⁷¹, Yongmei Liu³¹⁷, Nicholas G. Martin⁶¹, Winfried März^{51,164,318}, Mads Melbye^{309,319}, Andres Metspalu^{14,147}, Susanne Moebus⁵³, Patricia B. Munroe¹⁸¹, Inger Njølstad¹⁸⁵, Ben A. Oostra^{32,84,320}, Colin N. A. Palmer¹⁰⁶, Nancy L. Pedersen¹⁵, Markus Perola^{14,53,107}

Louis Pérusse^{278,302}, Ulrike Peters⁴⁹, Chris Power²⁵⁷, Thomas Quertermous³⁰⁹, Rainer Rauramaa^{125,273}, Fernando Rivadeneira^{58,59,60}, Timo E. Saaristo^{321,322}, Danish Saleheen^{232,323,324}, Naveed Sattar³²⁵, Eric E. Schadt³²⁶, David Schlessinger¹⁵⁵, P. Eline Slagboom^{34,35}, Harold Snieder¹⁶⁹, Tim D. Spector⁵⁶, Unnur Thorsteinsdottir^{171,327}, Michael Stumvoll^{96,97}, Jaakko Tuomilehto^{53,328,329,330}, André G. Uitterlinden^{58,59,60}, Matti Uusitupa^{331,332}, Pim van der Harst^{29,57,264}, Mark Walker³³³, Henri Wallaschowski^{239,334}, Nicholas J. Wareham⁹, Hugh Watkins^{18,48}, David R. Weir²⁶, H-Erich Wichmann^{335,336,337}, James F. Wilson³⁶, Pieter Zanen³³⁸, Ingrid B. Borecki⁴⁷, Panos Deloukas^{22,50,339}, Caroline S. Fox¹²⁷, Iris M. Heid^{23,85}, Jeffrey R. O'Connell^{295,296}, David P. Strachan³⁴⁰, Kari Stefansson^{171,327}, Cornelia M. van Duijn^{32,58,59,84}, Gonçalo R. Abecasis¹, Lude Franke²⁹, Timothy M. Frayling²⁴, Mark I. McCarthy^{18,64,341}, Peter M. Visscher^{11,12}, André Scherag^{63,342}, Cristen J. Willer^{28,30,343}, Michael Boehnke⁴, Karen L. Mohlke¹⁰, Cecilia M. Lindgren^{6,18}, Jacques S. Beckmann^{20,21,344}, Inês Barroso^{22,345,346}, Kari E. North^{4,347}, Erik Ingelsson^{16,17,18}, Joel N. Hirschhorn^{5,6,7}, Ruth J. F. Loos^{9,55,145,348}, & Elizabeth K. Speliotes²

¹Center for Statistical Genetics, Department of Biostatistics, University of Michigan, Ann Arbor, Michigan 48109, USA. ²Department of Internal Medicine, Division of Gastroenterology, and Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, Michigan 48109, USA. ³Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, USA. ⁴Department of Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA. ⁵Divisions of Endocrinology and Genetics and Center for Basic and Translational Obesity Research, Boston Children's Hospital, Boston, Massachusetts 02115, USA. ⁶Broad Institute of the Massachusetts Institute of Technology and Harvard University, Cambridge, Massachusetts 02142, USA. ⁷Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA. ⁸Center for Biological Sequence Analysis, Department of Systems Biology, Technical University of Denmark, Lyngby 2800, Denmark. ⁹MRC Epidemiology Unit, University of Cambridge School of Clinical Medicine, Institute of Metabolic Science, Cambridge Biomedical Campus, Cambridge CB2 0QQ, UK. ¹⁰Department of Genetics, University of North Carolina, Chapel Hill, North Carolina 27599, USA. ¹¹Queensland Brain Institute, The University of Queensland, Brisbane 4072, Australia. ¹²The University of Queensland Diamantina Institute, The Translation Research Institute, Brisbane 4012, Australia. ¹³Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts 02115, USA. ¹⁴Estonian Genome Center, University of Tartu, Tartu 51010, Estonia. ¹⁵Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm 17177, Sweden. ¹⁶Science for Life Laboratory, Uppsala University, Uppsala 75185, Sweden. ¹⁷Department of Medical Sciences, Molecular Epidemiology, Uppsala University, Uppsala 75185, Sweden. ¹⁸Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK. ¹⁹Institute of Social and Preventive Medicine (IUMSP), Centre Hospitalier Universitaire Vaudois (CHUV), Lausanne 1010, Switzerland. ²⁰Swiss Institute of Bioinformatics, Lausanne 1015, Switzerland. ²¹Department of Medical Genetics, University of Lausanne, Lausanne 1005, Switzerland. ²²Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1SA, UK. ²³Department of Genetic Epidemiology, Institute of Epidemiology and Preventive Medicine, University of Regensburg, D-93053 Regensburg, Germany. ²⁴Genetics of Complex Traits, University of Exeter Medical School, University of Exeter, Exeter EX1 2LU, UK. ²⁵Department of Nutrition, Harvard School of Public Health, Boston, Massachusetts 02115, USA. ²⁶Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, Michigan 48104, USA. ²⁷Department of Epidemiology, University of Michigan, Ann Arbor, Michigan 48109, USA. ²⁸Department of Internal Medicine, Division of Cardiovascular Medicine, University of Michigan, Ann Arbor, Michigan 48109, USA. ²⁹Department of Genetics, University Medical Center Groningen, University of Groningen, 9700 RB Groningen, The Netherlands. ³⁰Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, Michigan 48109, USA. ³¹HudsonAlpha Institute for Biotechnology, Huntsville, Alabama 35806, USA. ³²Genetic Epidemiology Unit, Department of Epidemiology, Erasmus MC University Medical Center, 3015 GE Rotterdam, The Netherlands. ³³Telethon Institute for Child Health Research, Centre for Child Health Research, The University of Western Australia, Perth, Western Australia 6008, Australia. ³⁴Netherlands Consortium for Healthy Aging (NCHA), Leiden University Medical Center, Leiden 2300 RC, The Netherlands. ³⁵Department of Molecular Epidemiology, Leiden University Medical Center, 2300 RC Leiden, The Netherlands. ³⁶Centre for Population Health Sciences, University of Edinburgh, Teviot Place, Edinburgh EH8 9AG, UK. ³⁷Kidney Epidemiology and Cost Center, University of Michigan, Ann Arbor, Michigan 48109, USA. ³⁸Department of Statistics & Biostatistics, Rutgers University, Piscataway, New Jersey 08854, USA. ³⁹Department of Genetics, Rutgers University, Piscataway, New Jersey 08854, USA. ⁴⁰Department of Human Genetics, Leiden University Medical Center, 2333 ZC Leiden, The Netherlands. ⁴¹Ealing Hospital NHS Trust, Middlesex UB1 3HW, UK. ⁴²Department of Gastroenterology and Hepatology, Imperial College London, London W2 1PG, UK. ⁴³Institute of Infectious Diseases, Southwest Hospital, Third Military Medical University, Chongqing, China. ⁴⁴Center for Complex Disease Genomics, McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA. ⁴⁵Cardiology, Department of Specialties of Internal Medicine, Geneva University Hospital, Geneva 1211, Switzerland. ⁴⁶Department of Epidemiology Research, Statens Serum Institut, Copenhagen DK-2300, Denmark. ⁴⁷Department of Genetics, Washington University School of Medicine, St Louis, Missouri 63110, USA. ⁴⁸Division of Cardiovascular Medicine, Radcliffe Department of Medicine, University of Oxford, Oxford OX3 9DU, UK. ⁴⁹Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington 98109, USA. ⁵⁰William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London EC1M 6BQ, UK. ⁵¹Vth Department of Medicine (Nephrology, Hypertensiology, Endocrinology, Diabetology, Rheumatology), Medical Faculty of Mannheim, University of Heidelberg, D-68187 Mannheim, Germany. ⁵²Department of

Internal Medicine II, Ulm University Medical Centre, D-89081 Ulm, Germany.⁵³National Institute for Health and Welfare, FI-00271 Helsinki, Finland.⁵⁴Epidemiology Program, University of Hawaii Cancer Center, Honolulu, Hawaii 96813, USA.⁵⁵The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA.⁵⁶Department of Twin Research and Genetic Epidemiology, King's College London, London SE1 7EH, UK.⁵⁷Department of Cardiology, University Medical Center Groningen, University of Groningen, 9700RB Groningen, The Netherlands.⁵⁸Netherlands Consortium for Healthy Aging (NCHA), 3015GE Rotterdam, The Netherlands.⁵⁹Department of Epidemiology, Erasmus MC University Medical Center, 3015GE Rotterdam, The Netherlands.⁶⁰Department of Internal Medicine, Erasmus MC University Medical Center, 3015GE Rotterdam, The Netherlands.⁶¹QIMR Berghofer Medical Research Institute, Brisbane, Queensland 4006, Australia.⁶²Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, Maryland 20892, USA.⁶³Institute for Medical Informatics, Biometry and Epidemiology (IMIBE), University Hospital Essen, 45147 Essen, Germany.⁶⁴Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Oxford OX3 7LJ, UK.⁶⁵Department of Genomics of Common Disease, School of Public Health, Imperial College London, Hammersmith Hospital, London W12 0NN, UK.⁶⁶Department of Clinical Sciences, Genetic & Molecular Epidemiology Unit, Lund University Diabetes Center, Skåne University Hospital, Malmö 205 02, Sweden.⁶⁷Department of Public Health and Clinical Medicine, Unit of Medicine, Umeå University, Umeå 901 87, Sweden.⁶⁸Department of Odontology, Umeå University, Umeå 901 85, Sweden.⁶⁹University of Eastern Finland, FI-70210 Kuopio, Finland.⁷⁰Atherosclerosis Research Unit, Center for Molecular Medicine, Department of Medicine, Karolinska Institutet, Stockholm 17176, Sweden.⁷¹Division of Biostatistics, Washington University School of Medicine, St Louis, Missouri 63110, USA.⁷²Translational Gerontology Branch, National Institute on Aging, Baltimore, Maryland 21225, USA.⁷³Interfaculty Institute for Genetics and Functional Genomics, University Medicine Greifswald, D-17475 Greifswald, Germany.⁷⁴Department of Cardiology, Leiden University Medical Center, 2300 RC Leiden, The Netherlands.⁷⁵Department of Gerontology and Geriatrics, Leiden University Medical Center, 2300 RC Leiden, The Netherlands.⁷⁶Experimental Cardiology Laboratory, Division Heart and Lungs, University Medical Center Utrecht, 3584 CX Utrecht, The Netherlands.⁷⁷Department of Medical Genetics, University Medical Center Utrecht, 3584 CX Utrecht, The Netherlands.⁷⁸Department of Endocrinology, University of Groningen, University Medical Center Groningen, 9700 RB Groningen, The Netherlands.⁷⁹Core Genotyping Facility, SAIC-Frederick, Inc., NCI-Frederick, Frederick, Maryland 21702, USA.⁸⁰CNRS UMR 8199, F-59019 Lille, France.⁸¹European Genomic Institute for Diabetes, F-59000 Lille, France.⁸²Université de Lille 2, F-59000 Lille, France.⁸³Department of Epidemiology and Biostatistics, Imperial College London, London W2 1PG, UK.⁸⁴Center for Medical Systems Biology, 2300 RC Leiden, The Netherlands.⁸⁵Institute of Genetic Epidemiology, Helmholtz Zentrum München - German Research Center for Environmental Health, D-85764 Neuherberg, Germany.⁸⁶School of Health and Social Studies, Dalarna University, SE-791 88 Falun, Sweden.⁸⁷PathWest Laboratory Medicine of Western Australia, Nedlands, Western Australia 6009, Australia.⁸⁸Department of Haematology, University of Cambridge, Cambridge CB2 0PT, UK.⁸⁹NHS Blood and Transplant, Cambridge CB2 0PT, UK.⁹⁰Geriatric Unit, Azienda Sanitaria Firenze (ASF), 50125 Florence, Italy.⁹¹USC-Office of Population Studies Foundation, Inc., University of San Carlos, Cebu City 6000, Philippines.⁹²Department of Genetics, Texas Biomedical Research Institute, San Antonio, Texas 78227, USA.⁹³Genomics Research Centre, Institute of Health and Biomedical Innovation, Queensland University of Technology, Brisbane, Queensland 4001, Australia.⁹⁴Department of Medical Sciences, Endocrinology, Diabetes and Metabolism, Uppsala University, Uppsala 75185, Sweden.⁹⁵Division of Endocrinology, Diabetes and Metabolism, Ulm University Medical Centre, D-89081 Ulm, Germany.⁹⁶Integrated Research and Treatment Center (IFB) Adiposity Diseases, University of Leipzig, D-04103 Leipzig, Germany.⁹⁷Department of Medicine, University of Leipzig, D-04103 Leipzig, Germany.⁹⁸Department of Medical Statistics and Bioinformatics, Leiden University Medical Center, 2300 RC Leiden, The Netherlands.⁹⁹Medical Genomics and Metabolic Genetics Branch, National Human Genome Research Institute, NIH, Bethesda, Maryland 20892, USA.¹⁰⁰LifeLines Cohort Study, University Medical Center Groningen, University of Groningen, 9700 RB Groningen, The Netherlands.¹⁰¹Department of Biology, Norwegian University of Science and Technology, 7491 Trondheim, Norway.¹⁰²Department of Pediatrics, University of California Los Angeles, Torrance, California 90502, USA.¹⁰³Transgenomics Institute, Los Angeles Biomedical Research Institute, Torrance, California 90502, USA.¹⁰⁴Clinical Trial Service Unit and Epidemiological Studies Unit, Nuffield Department of Population Health, University of Oxford, Oxford OX3 7LF, UK.¹⁰⁵Department of Dietetics-Nutrition, Harokopio University, 17671 Athens, Greece.¹⁰⁶Medical Research Institute, University of Dundee, Ninewells Hospital and Medical School, Dundee DD1 9SY, UK.¹⁰⁷Institute for Molecular Medicine, University of Helsinki, FI-00014 Helsinki, Finland.¹⁰⁸Analytic and Translational Genetics Unit, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts 02114, USA.¹⁰⁹Laboratory of Epidemiology and Population Sciences, National Institute on Aging, NIH, Bethesda, Maryland 20892, USA.¹¹⁰Department of Public Health and Caring Sciences, Geriatrics, Uppsala University, Uppsala 75185, Sweden.¹¹¹Division of Cardiovascular Epidemiology, Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden, Stockholm 17177, Sweden.¹¹²Kaiser Permanente, Division of Research, Oakland, California 94612, USA.¹¹³Service of Therapeutic Education for Diabetes, Obesity and Chronic Diseases, Geneva University Hospital, Geneva CH-1211, Switzerland.¹¹⁴Department of Cardiovascular Sciences, University of Leicester, Glenfield Hospital, Leicester LE3 9QP, UK.¹¹⁵National Institute for Health Research (NIHR) Leicester Cardiovascular Biomedical Research Unit, Glenfield Hospital, Leicester LE3 9QP, UK.¹¹⁶Department of Nephrology, University Hospital Regensburg, D-93053 Regensburg, Germany.¹¹⁷Department of Psychiatry and Psychotherapy, University Medicine Greifswald, HELIOS-Hospital Stralsund, D-17475 Greifswald, Germany.¹¹⁸German Center for Neurodegenerative Diseases (DZNE), Rostock, Greifswald, D-17475 Greifswald, Germany.¹¹⁹Research Unit of Molecular Epidemiology, Helmholtz Zentrum München - German Research Center for Environmental Health, D-85764 Neuherberg, Germany.¹²⁰German Center for Diabetes Research (DZD), 85764 Neuherberg, Germany.¹²¹Department of Medicine III, University Hospital Carl Gustav Carus, Technische Universität Dresden, D-01307 Dresden, Germany.¹²²Institut inter Régional pour la Santé, Synergies, F-37520 La Riche, France.¹²³Department of Public Health and Clinical Medicine, Unit of Nutritional Research, Umeå University, Umeå 90187, Sweden.¹²⁴Department of Psychiatry, University of Groningen, University Medical Center Groningen, 9700RB Groningen, The Netherlands.¹²⁵Kuopio Research Institute of Exercise Medicine, 70100 Kuopio, Finland.¹²⁶MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Western General Hospital, Edinburgh EH4 2XU, UK.¹²⁷National Heart, Lung, and Blood Institute, the Framingham Heart Study, Framingham, Massachusetts 01702, USA.¹²⁸Department of Neurology, Boston University School of Medicine, Boston, Massachusetts 02118, USA.¹²⁹Faculty of Psychology and Education, VU University Amsterdam, 1081BT Amsterdam, The Netherlands.¹³⁰Deutsches Forschungszentrum für Herz-Kreislauferkrankungen (DZHK) (German Research Centre for Cardiovascular Research), Munich Heart Alliance, D-80636 Munich, Germany.¹³¹Deutsches Herzzentrum München, Technische Universität München, D-80636 Munich, Germany.¹³²Department of Public Health and General Practice, Norwegian University of Science and Technology, Trondheim 7489, Norway.¹³³Biological Psychology, VU University Amsterdam, 1081BT Amsterdam, The Netherlands.¹³⁴Department of Pulmonary Physiology and Sleep Medicine, Nedlands, Western Australia 6009, Australia.¹³⁵School of Medicine and Pharmacology, University of Western Australia, Crawley 6009, Australia.¹³⁶Uppsala University, Department of Immunology, Genetics, Pathology, SciLifeLab, Rudbeck Laboratory, SE-751 85 Uppsala, Sweden.¹³⁷Hjelt Institute Department of Public Health, University of Helsinki, FI-00014 Helsinki, Finland.¹³⁸Department of Internal Medicine I, Ulm University Medical Centre, D-89081 Ulm, Germany.¹³⁹Finnish Institute of Occupational Health, FI-90100 Oulu, Finland.¹⁴⁰Division of Genetic Epidemiology, Department of Medical Genetics, Molecular and Clinical Pharmacology, Innsbruck Medical University, 6020 Innsbruck, Austria.¹⁴¹Institute of Human Genetics, Helmholtz Zentrum München - German Research Center for Environmental Health, D-85764 Neuherberg, Germany.¹⁴²Department of Medical Sciences, Cardiovascular Epidemiology, Uppsala University, Uppsala 75185, Sweden.¹⁴³Montreal Heart Institute, Montreal, Quebec H1T 1C8, Canada.¹⁴⁴Institute for Community Medicine, University Medicine Greifswald, D-17475 Greifswald, Germany.¹⁴⁵The Genetics of Obesity and Related Metabolic Traits Program, The Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA.¹⁴⁶School of Social and Community Medicine, University of Bristol, Bristol BS8 2BN, UK.¹⁴⁷Institute of Molecular and Cell Biology, University of Tartu, Tartu 51010, Estonia.¹⁴⁸Farr Institute of Health Informatics Research, University College London, London NW1 2DA, UK.¹⁴⁹The Center for Observational Research, Amgen, Inc., Thousand Oaks, California 91320, USA.¹⁵⁰Istituto di Ricerca Genetica e Biomedica (IRGB), Consiglio Nazionale delle Ricerche, Cagliari, Sardinia 09042, Italy.¹⁵¹Center for Evidence-based Healthcare, University Hospital Carl Gustav Carus, Technische Universität Dresden, D-01307 Dresden, Germany.¹⁵²Department of Medicine I, University Hospital Grosshadern, Ludwig-Maximilians-Universität, D-81377 Munich, Germany.¹⁵³Institute of Medical Informatics, Biometry and Epidemiology, Chair of Genetic Epidemiology, Ludwig-Maximilians-Universität, D-81377 Munich, Germany.¹⁵⁴Department of Respiratory Medicine, Sir Charles Gairdner Hospital, Nedlands, Western Australia 6009, Australia.¹⁵⁵Laboratory of Genetics, National Institute on Aging, Baltimore, Maryland 21224, USA.¹⁵⁶Department of Genomics, Life & Brain Center, University of Bonn, 53127 Bonn, Germany.¹⁵⁷Institute of Human Genetics, University of Bonn, 53127 Bonn, Germany.¹⁵⁸Department of Epidemiology, University Medical Center Groningen, University of Groningen, 9700 RB Groningen, The Netherlands.¹⁵⁹Department of Epidemiology and Biostatistics, Institute for Research in Extramural Medicine, Institute for Health and Care Research, VU University Medical Center, 1081BT Amsterdam, The Netherlands.¹⁶⁰Department of Internal Medicine, Division of Endocrinology and Metabolism, Medical University of Graz, 8036 Graz, Austria.¹⁶¹Institute of Physiology, University Medicine Greifswald, D-17495 Karlsburg, Germany.¹⁶²Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA.¹⁶³Division of Preventive Medicine, Brigham and Women's Hospital, Boston, Massachusetts 02215, USA.¹⁶⁴Clinical Institute of Medical and Chemical Laboratory Diagnostics, Medical University of Graz, Graz 8036, Austria.¹⁶⁵Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, California 90089, USA.¹⁶⁶National Cancer Institute, Bethesda, Maryland 20892, USA.¹⁶⁷Icelandic Heart Association, Kopavogur 201, Iceland.¹⁶⁸University of Iceland, Reykjavik 101, Iceland.¹⁶⁹Department of Epidemiology, University Medical Center Groningen, University of Groningen, 9700 RB Groningen, The Netherlands.¹⁷⁰Molecular & Cellular Therapeutics, Royal College of Surgeons in Ireland, 123 St Stephen's Green, Dublin 2, Ireland.¹⁷¹deCODE Genetics, Amgen Inc., Reykjavik 101, Iceland.¹⁷²Department of Medical Sciences, Molecular Medicine, Uppsala University, Uppsala 75144, Sweden.¹⁷³National Heart and Lung Institute, Imperial College London, London SW3 6LY, UK.¹⁷⁴Department of Public Health Sciences, Stritch School of Medicine, Loyola University of Chicago, Maywood, Illinois 61053, USA.¹⁷⁵Institute of Epidemiology II, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany, D-85764 Neuherberg, Germany.¹⁷⁶Department of Oncology, University of Cambridge, Cambridge CB2 0QQ, UK.¹⁷⁷Centre for Bone and Arthritis Research, Department of Internal Medicine and Clinical Nutrition, Institute of Medicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg 413 45, Sweden.¹⁷⁸Department of Child and Adolescent Psychiatry/Psychology, Erasmus MC University Medical Centre, 3000 CB Rotterdam, The Netherlands.¹⁷⁹Department for Health Evidence, Radboud University Medical Centre, 6500 HB Nijmegen, The Netherlands.¹⁸⁰Department of Genetics, Radboud University Medical Centre, 6500 HB Nijmegen, The Netherlands.¹⁸¹Department of Clinical Pharmacology, William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London EC1M 6BQ, UK.¹⁸²Genetics, GlaxoSmithKline, King of Prussia, Pennsylvania 19406, USA.¹⁸³German Center for Cardiovascular Research, partner site Hamburg/Lubeck/Kiel, 23562 Lubeck, Germany.¹⁸⁴Institut für Integrative und Experimentelle Genomik, Universität zu Lubeck, D-23562 Lubeck, Germany.¹⁸⁵Department of Community Medicine, Faculty of Health Sciences, UiT The Arctic

University of Norway, 9037 Tromsø, Norway.¹⁸⁶MRC Unit for Lifelong Health and Ageing at University College London, London WC1B 5JU, UK.¹⁸⁷Diabetes Complications Research Centre, Conway Institute, School of Medicine and Medical Sciences, University College Dublin, Dublin 4, Ireland.¹⁸⁸Department of Biomedical Sciences, Seoul National University College of Medicine, Seoul, Korea.¹⁸⁹Lady Davis Institute, Departments of Human Genetics, Epidemiology and Biostatistics, McGill University, Montréal, Québec H3T1E2, Canada.¹⁹⁰Cardiothoracic Surgery Unit, Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm 17176, Sweden.¹⁹¹Department of Medicine, Columbia University College of Physicians and Surgeons, New York 10032, USA.¹⁹²Biosciences Research Division, Department of Primary Industries, Victoria 3083, Australia.¹⁹³Department of Food and Agricultural Systems, University of Melbourne, Victoria 3010, Australia.¹⁹⁴Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts 02115, USA.¹⁹⁵State Key Laboratory of Medical Genomics, Shanghai Institute of Hematology, Rui Jin Hospital Affiliated with Shanghai Jiao Tong University School of Medicine, Shanghai, China.¹⁹⁶NIHR Oxford Biomedical Research Centre, OUH Trust, Oxford OX3 7LE, UK.¹⁹⁷Cardiovascular Research Center, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts, USA.¹⁹⁸Laboratory for Genotyping Development, RIKEN Center for Integrative Medical Sciences, Yokohama 230-0045, Japan.¹⁹⁹Center for Genome Science, National Institute of Health, Chungcheongbuk-do, Chungbuk 363-951, Republic of Korea.²⁰⁰Harvard School of Public Health, Department of Biostatistics, Harvard University, Boston, Massachusetts 2115, USA.²⁰¹Department of Genetics, Howard Hughes Medical Institute, Yale University School of Medicine, New Haven, New Haven, Connecticut 06520, USA.²⁰²College of Information Science and Technology, Dalian Maritime University, Dalian, Liaoning 116026, China.²⁰³Nephrology Research, Centre for Public Health, Queen's University of Belfast, Belfast, County Down BT9 7AB, UK.²⁰⁴Section of General Internal Medicine, Boston University School of Medicine, Boston, Massachusetts 02118, USA.²⁰⁵Department of Statistics, University of Oxford, 1 South Parks Road, Oxford OX1 3TG, UK.²⁰⁶MRC Harwell, Harwell Science and Innovation Campus, Harwell OX11 0QG, UK.²⁰⁷Institute of Health and Biomedical Innovation, Queensland University of Technology, Brisbane, Queensland 4059, Australia.²⁰⁸Laboratory for Statistical Analysis, RIKEN Center for Integrative Medical Sciences, Yokohama 230-0045, Japan.²⁰⁹Department of Human Genetics and Disease Diversity, Graduate School of Medical and Dental Sciences, Tokyo Medical and Dental University, 113-8510 Tokyo, Japan.²¹⁰Genome Institute of Singapore, Agency for Science, Technology and Research, 138672 Singapore.²¹¹Department of Biomedical Engineering and Computational Science, Aalto University School of Science, Helsinki FI-00076, Finland.²¹²Department of Medicine, Division of Nephrology, Helsinki University Central Hospital, FI-00290 Helsinki, Finland.²¹³Folkhälsan Institute of Genetics, Folkhälsan Research Center, FI-00290 Helsinki, Finland.²¹⁴Laboratory for Cardiovascular Diseases, RIKEN Center for Integrative Medical Sciences, Yokohama 230-0045, Japan.²¹⁵Division of Disease Diversity, Bioresearch Research Center, Tokyo Medical and Dental University, 113-8510 Tokyo, Japan.²¹⁶Division of Epidemiology, Department of Medicine; Vanderbilt Epidemiology Center; and Vanderbilt-Ingram Cancer Center, Vanderbilt University Medical Center, Nashville, Tennessee 37075, USA.²¹⁷Nuffield Department of Obstetrics & Gynaecology, University of Oxford, Oxford OX3 7BN, UK.²¹⁸Department of Psychiatry, Washington University School of Medicine, St Louis, Missouri 63110, USA.²¹⁹Department of Epidemiology and Public Health, EA3430, University of Strasbourg, Faculty of Medicine, Strasbourg, France.²²⁰Department of Internal Medicine, University Medical Center Groningen, University of Groningen, 9700RB Groningen, The Netherlands.²²¹Pathology and Laboratory Medicine, The University of Western Australia, Perth, Western Australia 6009, Australia.²²²Cedars-Sinai Diabetes and Obesity Research Institute, Los Angeles, California 90048, USA.²²³Institute of Social and Preventive Medicine (IUMSP), Centre Hospitalier Universitaire Vaudois and University of Lausanne, 1010 Lausanne, Switzerland.²²⁴Ministry of Health, Victoria, Republic of Seychelles.²²⁵University of Milano, Bicocca, 20126, Italy.²²⁶Harvard Medical School, Boston, Massachusetts 02115, USA.²²⁷Center for Human Genetics Research, Vanderbilt University Medical Center, Nashville, Tennessee 37203, USA.²²⁸Department of Molecular Physiology and Biophysics, Vanderbilt University, Nashville, Tennessee 37232, USA.²²⁹Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts 02118, USA.²³⁰Department of Health Sciences, University of Milano, I 20142, Italy.²³¹Fondazione Filarete, Milano I 20139, Italy.²³²Department of Public Health and Primary Care, University of Cambridge, Cambridge CB1 8RN, UK.²³³Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, 3584 CX Utrecht, The Netherlands.²³⁴Institute of Cardiovascular and Medical Sciences, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow G12 8TA, UK.²³⁵Clinic of Cardiology, West-German Heart Centre, University Hospital Essen, 45147 Essen, Germany.²³⁶Department of General Practice and Primary Health Care, University of Helsinki, FI-00290 Helsinki, Finland.²³⁷Unit of General Practice, Helsinki University Central Hospital, Helsinki 00290, Finland.²³⁸Department of Internal Medicine B, University Medicine Greifswald, D-17475 Greifswald, Germany.²³⁹DZHK (Deutsches Zentrum für Herz-Kreislauforschung – German Centre for Cardiovascular Research), partner site Greifswald, D-17475 Greifswald, Germany.²⁴⁰Department of Internal Medicine, University of Pisa, 56100 Pisa, Italy.²⁴¹National Research Council Institute of Clinical Physiology, University of Pisa, 56124 Pisa, Italy.²⁴²Department of Cardiology, Toulouse University School of Medicine, Rangueil Hospital, 31400 Toulouse, France.²⁴³Robertson Center for Biostatistics, University of Glasgow, Glasgow G12 8QQ, UK.²⁴⁴UWI Solutions for Developing Countries, The University of the West Indies, Mona, Kingston 7, Jamaica.²⁴⁵NorthShore University HealthSystem, Evanston, IL 60201, University of Chicago, Chicago, Illinois, USA.²⁴⁶Leeds MRC Medical Bioinformatics Centre, University of Leeds, Leeds LS2 9LU, UK.²⁴⁷Institute of Biomedical & Clinical Science, University of Exeter, Barrack Road, Exeter EX2 5DW, UK.²⁴⁸Center for Biomedicine, European Academy Bozen, Bolzano (EURAC), Bolzano 39100, Italy (affiliated institute of the University of Lübeck, D-23562 Lübeck, Germany).²⁴⁹Division of Genomic Medicine, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland 20892, USA.²⁵⁰Institute of Cardiovascular Science, University College London, London WC1E 6BT, UK.²⁵¹Department of Vascular Medicine, Academic Medical Center, 1105 AZ Amsterdam, The Netherlands.²⁵²Centre for Cardiovascular Genetics, Institute Cardiovascular Sciences, University College London, London WC1E 6JJ, UK.²⁵³Cardiovascular Genetics Division, Department of Internal Medicine, University of Utah, Salt Lake City, Utah 84108, USA.²⁵⁴Sansom Institute for Health Research, University of South Australia, Adelaide 5000, South Australia, Australia.²⁵⁵School of Population Health, University of South Australia, Adelaide 5000, South Australia, Australia.²⁵⁶South Australian Health and Medical Research Institute, Adelaide, South Australia 5000, Australia.²⁵⁷Population, Policy, and Practice, University College London Institute of Child Health, London WC1N 1EH, UK.²⁵⁸Hannover Unified Biobank, Hannover Medical School, Hannover, D-30625 Hannover, Germany.²⁵⁹National Institute for Health and Welfare, FI-90101 Oulu, Finland.²⁶⁰MRC Health Protection Agency (HPA) Centre for Environment and Health, School of Public Health, Imperial College London, London W2 1PG, UK.²⁶¹Unit of Primary Care, Oulu University Hospital, FI-90220 Oulu, Finland.²⁶²Biocenter Oulu, University of Oulu, FI-90014 Oulu, Finland.²⁶³Institute of Health Sciences, University of Oulu, FI-90014 Oulu, Finland.²⁶⁴Durrer Center for Cardiogenetic Research, Interuniversity Cardiology Institute Netherlands (ICIN), 3501 DG Utrecht, The Netherlands.²⁶⁵Interuniversity Cardiology Institute of the Netherlands (ICIN), 3501 DG Utrecht, The Netherlands.²⁶⁶Unit of Primary Health Care/General Practice, Oulu University Hospital, FI-90220 Oulu, Finland.²⁶⁷Department of Urology, Radboud University Medical Centre, 6500 HB Nijmegen, The Netherlands.²⁶⁸Imperial College Healthcare NHS Trust, London W12 0HS, UK.²⁶⁹Department of Epidemiology and Public Health, University College London, London WC1E 6BT, UK.²⁷⁰Department of Biological and Social Epidemiology, University of Essex, Wivenhoe Park, Colchester, Essex CO4 3SQ, UK.²⁷¹Department of Medicine, Kuopio University Hospital and University of Eastern Finland, FI-70210 Kuopio, Finland.²⁷²Department of Physiology, Institute of Biomedicine, University of Eastern Finland, Kuopio Campus, FI-70211 Kuopio, Finland.²⁷³Department of Clinical Physiology and Nuclear Medicine, Kuopio University Hospital and University of Eastern Finland, FI-70210 Kuopio, Finland.²⁷⁴Department of Clinical Chemistry, Fimlab Laboratories and School of Medicine University of Tampere, FI-33520 Tampere, Finland.²⁷⁵Steno Diabetes Center A/S, Gentofte DK-2820, Denmark.²⁷⁶Lund University Diabetes Centre and Department of Clinical Science, Diabetes & Endocrinology Unit, Lund University, Malmö 221 00, Sweden.²⁷⁷Institut Universitaire de Cardiologie et de Pneumologie de Québec, Faculty of Medicine, Laval University, Québec, QC G1V 0A6, Canada.²⁷⁸Institute of Nutrition and Functional Foods, Laval University, Québec, QC G1V 0A6, Canada.²⁷⁹Department of Biostatistics, University of Washington, Seattle, Washington 98195, USA.²⁸⁰Department of Surgery, University Medical Center Utrecht, 3584 CX Utrecht, The Netherlands.²⁸¹Department of Biostatistics, University of Liverpool, Liverpool L69 3GA, UK.²⁸²Department of Pediatrics, University of Iowa, Iowa City, Iowa 52242, USA.²⁸³illumina, Inc, Little Chesterford, Cambridge CB10 1XL, UK.²⁸⁴University of Groningen, University Medical Center Groningen, Department of Pulmonary Medicine and Tuberculosis, Groningen, The Netherlands.²⁸⁵Department of Neurology, General Central Hospital, Bolzano 39100, Italy.²⁸⁶Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, FI-20521 Turku, Finland.²⁸⁷Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, FI-20521 Turku, Finland.²⁸⁸Human Genomics Laboratory, Pennington Biomedical Research Center, Baton Rouge, Louisiana 70808, USA.²⁸⁹Université de Montréal, Montreal, Quebec H1T 1C8, Canada.²⁹⁰Center for Systems Genomics, The Pennsylvania State University, University Park, Pennsylvania 16802, USA.²⁹¹Croatian Centre for Global Health, Faculty of Medicine, University of Split, 21000 Split, Croatia.²⁹²South Carelia Central Hospital, 53130 Lappeenranta, Finland.²⁹³Paul Langerhans Institute Dresden, German Center for Diabetes Research (DZD), 01307 Dresden, Germany.²⁹⁴International Centre for Circulatory Health, Imperial College London, London W2 1PG, UK.²⁹⁵Division of Endocrinology, Diabetes and Nutrition, University of Maryland School of Medicine, Baltimore, Maryland 21201, USA.²⁹⁶Program for Personalized and Genomic Medicine, University of Maryland School of Medicine, Baltimore, Maryland 21201, USA.²⁹⁷Geriatric Research and Education Clinical Center, Veterans Administration Medical Center, Baltimore, Maryland 21201, USA.²⁹⁸Helsinki University Central Hospital Heart and Lung Center, Department of Medicine, Helsinki University Central Hospital, FI-00290 Helsinki, Finland.²⁹⁹Sorbonne Universités, UPMC Univ Paris 06, UMR S 1166, F-75013 Paris, France.³⁰⁰INSERM, UMR S 1166, Team Genomics and Physiopathology of Cardiovascular Diseases, F-75013 Paris, France.³⁰¹Institute for Cardiometabolism And Nutrition (ICAN), F-75013 Paris, France.³⁰²Department of Kinesiology, Laval University, Québec QC G1V 0A6, Canada.³⁰³Dipartimento di Scienze Farmacologiche e Biomolecolari, Università di Milano & Centro Cardiologico Monzino, Istituto di Ricovero e Cura a Carattere Scientifico, Milan 20133, Italy.³⁰⁴Department of Food Science and Nutrition, Laval University, Québec QC G1V 0A6, Canada.³⁰⁵Department of Internal Medicine, University Hospital (CHUV) and University of Lausanne, Lausanne 1011, Switzerland.³⁰⁶Department of Nutrition, University of North Carolina, Chapel Hill, North Carolina 27599, USA.³⁰⁷Institut Pasteur de Lille; INSERM, U744; Université de Lille 2; F-59000 Lille, France.³⁰⁸Department of Cardiology, Division Heart and Lungs, University Medical Center Utrecht, 3584 CX Utrecht, The Netherlands.³⁰⁹Department of Medicine, Stanford University School of Medicine, Palo Alto, California 94304, USA.³¹⁰Lee Kong Chian School of Medicine, Imperial College London and Nanyang Technological University, Singapore, 637553 Singapore, Singapore.³¹¹Department of Internal Medicine I, Ulm University Medical Center, D-89081 Ulm, Germany.³¹²Health Science Center at Houston, University of Texas, Houston, Texas 77030, USA.³¹³Department of Medicine, Division of Genetics, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA.³¹⁴Department of Epidemiology, University Medical Center Utrecht, 3584 CX Utrecht, The Netherlands.³¹⁵School of Population Health, The University of Western Australia, Nedlands, Western Australia 6009, Australia.³¹⁶Albert Einstein College of Medicine, Department of Epidemiology and Population Health, Belfer 1306, New York 10461, USA.³¹⁷Center for Human Genetics, Division of Public Health Sciences, Wake Forest School of Medicine, Winston-Salem, North Carolina 27157, USA.³¹⁸Synlab Academy, Synlab Services GmbH, 68163 Mannheim, Germany.³¹⁹Department of Clinical Medicine, Copenhagen University, 2200 Copenhagen, Denmark.³²⁰Department of Clinical Genetics, Erasmus MC University Medical Center, 3000 CA Rotterdam, The Netherlands.³²¹Finnish Diabetes Association, Kirjoniementie 15, FI-33680 Tampere, Finland.³²²Pirkanmaa Hospital District, FI-33521 Tampere,

Finland.³²³Center for Non-Communicable Diseases, Karachi, Pakistan.³²⁴Department of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA.³²⁵BHF Glasgow Cardiovascular Research Centre, Division of Cardiovascular and Medical Sciences, University of Glasgow, Glasgow G12 8TA, UK.³²⁶Icahn Institute for Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, New York 10580, USA.³²⁷Faculty of Medicine, University of Iceland, Reykjavik 101, Iceland.³²⁸Institute for Health Research, University Hospital of La Paz (IdiPaz), 28046 Madrid, Spain.³²⁹Diabetes Research Group, King Abdulaziz University, 21589 Jeddah, Saudi Arabia.³³⁰Centre for Vascular Prevention, Danube-University Krems, 3500 Krems, Austria.³³¹Department of Public Health and Clinical Nutrition, University of Eastern Finland, Finland.³³²Research Unit, Kuopio University Hospital, FI-70210 Kuopio, Finland.³³³Institute of Cellular Medicine, Newcastle University, Newcastle NE1 7RU, UK.³³⁴Institute of Clinical Chemistry and Laboratory Medicine, University Medicine Greifswald, D-17475 Greifswald, Germany.³³⁵Institute of Medical Informatics, Biometry and Epidemiology, Chair of Epidemiology, Ludwig-Maximilians-Universität, D-85764 Munich, Germany.³³⁶Klinikum Grosshadern, D-81377 Munich, Germany.³³⁷Institute of Epidemiology I, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany, D-85764 Neuherberg, Germany.³³⁸Department of Pulmonology, University Medical Center Utrecht, 3584 CX Utrecht, The Netherlands.³³⁹Princess Al-Jawhara Al-Brahim Centre of Excellence in Research of

Hereditary Disorders (PACER-HD), King Abdulaziz University, 21589 Jeddah, Saudi Arabia.³⁴⁰Division of Population Health Sciences & Education, St George's, University of London, London SW17 0RE, UK.³⁴¹Oxford NIHR Biomedical Research Centre, Oxford University Hospitals NHS Trust, Oxford OX3 7LJ, UK.³⁴²Clinical Epidemiology, Integrated Research and Treatment Center, Center for Sepsis Control and Care (CSCC), Jena University Hospital, 07743 Jena, Germany.³⁴³Department of Human Genetics, University of Michigan, Ann Arbor, Michigan 48109, USA.³⁴⁴Service of Medical Genetics, CHUV University Hospital, 1011 Lausanne, Switzerland.³⁴⁵University of Cambridge Metabolic Research Laboratories, Institute of Metabolic Science, Addenbrooke's Hospital, Cambridge CB2 0QQ, UK.³⁴⁶NIHR Cambridge Biomedical Research Centre, Institute of Metabolic Science, Addenbrooke's Hospital, Cambridge CB2 0QQ, UK.³⁴⁷Carolina Center for Genome Sciences, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA.³⁴⁸The Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA.

†Present address: Second Floor, B-dong, AICT Building, 145 Gwanggyo-ro, Yeongyong-gu, Suwon-si, Gyeonggi-do, 443-270, South Korea.

‡A list of authors and affiliations appears in the Supplementary Information.

*These authors contributed equally to this work.

§These authors jointly supervised this work.

METHODS

Study design. We conducted a two-stage meta-analysis to identify BMI-associated loci in European adults (Extended Data Fig. 1 and Extended Data Table 1). In stage 1 we performed meta-analysis of 80 GWAS ($n = 234,069$); and stage 2 incorporated data from 34 additional studies ($n = 88,137$) genotyped using Metabochip⁷ (Supplementary Tables 1–3). Secondary meta-analyses were also conducted for: (1) all ancestries, (2) European men, (3) European women, and (4) European population-based studies. The total number of subjects and SNPs included in each stage for all analyses is shown in Extended Data Table 1. No statistical methods were used to predetermine sample size.

Phenotype. BMI, measured or self-reported weight in kg per height in metres squared (Supplementary Tables 1 and 3) was adjusted for age, age squared, and any necessary study-specific covariates (for example, genotype-derived principal components) in a linear regression model. The resulting residuals were transformed to approximate normality using inverse normal scores. For studies with no known related individuals, residuals were calculated separately by sex and case/control status. For family-based studies, residuals were calculated with men and women together, adding sex as an additional covariate in the linear regression model. Relatedness was accounted for in a study-specific manner (Supplementary Table 2).

Sample quality control, imputation and association. Following study-specific quality control measures (Supplementary Table 2), all contributing GWAS common SNPs were imputed using the HapMap phase II CEU reference panel for European-descent studies³⁷, and CEU+YRI+CHB+JPT HapMap release 22 for the African-American and Hispanic GWAS. Directly genotyped (GWAS and Metabochip) and imputed variants (GWAS only) were then tested for association with the inverse normally transformed BMI residuals using linear regression assuming an additive genetic model. Quality control following study level analyses was conducted following procedures outlined elsewhere³⁸.

Meta-analysis. Fixed effects meta-analyses were conducted using the inverse variance-weighted method implemented in METAL³⁹. Study-specific GWAS results as well as GWAS meta-analysis results were corrected for genomic control using all SNPs⁴⁰. Study-specific Metabochip results as well as Metabochip meta-analysis results were genomic-control-corrected using 4,425 SNPs included on Metabochip for replication of associations with QT-interval, a phenotype not correlated with BMI, after pruning of SNPs within 500 kb of an anthropometry replication SNP. The final meta-analysis combined the genomic-control-corrected GWAS and Metabochip meta-analysis results.

Identification of novel loci. We used a distance criterion of ± 500 kb surrounding each GWS peak ($P < 5 \times 10^{-8}$) to define independent loci and to place our results in the context of previous studies, including our previous GIANT meta-analyses. Of several locus models tested, this definition most closely reflected the loci defined by approximate conditional analysis using GCTA (Tables 1 and 2, respectively). Current index SNPs falling within 500 kb of a SNP previously associated with BMI, weight, extreme obesity or body fat percentage^{5,8–11} were considered previously identified.

Characterization of BMI-associated SNP effects. To investigate potential sources of heterogeneity between groups we compared the effect estimates of our 97 GWS SNPs for men versus women of European ancestry and Europeans versus non-Europeans. To address the effects of studies ascertained on a specific disease or phenotype on our results we also compare the effect estimates of European ancestry studies of population-based studies with the following European-descent subsets of studies: (1) non-population-based studies (that is, those ascertained on a specific disease or phenotype); (2) type 2 diabetes cases; (3) type 2 diabetes controls; (4) combined type 2 diabetes cases and controls; (5) CAD cases; (6) CAD controls; and (7) combined CAD cases and controls (Supplementary Tables 10 and 11). We also tested for heterogeneity of effect estimates between our European sex-combined meta-analysis and results from recent GWAS meta-analyses for BMI in individuals of African or east Asian ancestry^{10,41} (Supplementary Table 9). Heterogeneity was assessed as described previously⁴². A Bonferroni-corrected $P < 5 \times 10^{-4}$ (corrected for 97 tests) was used to assess significance. For heterogeneity tests assessing effects of ascertainment, we also used a 5% FDR threshold to assess significance of heterogeneity statistics (Supplementary Table 11).

Fine-mapping. We compared the meta-analysis results and credible sets of SNPs likely to contain the causal variant, based on the method described previously¹⁴, across the European-only, non-European, and all ancestries sex-combined meta-analyses. For each index SNP falling within a Metabochip fine-mapping region (27 for BMI), all SNPs available within 500 kb on either side of the index SNP were selected. Effect size estimates and standard errors for each SNP were converted to approximate Bayes' factors according to the method described previously¹⁵. All approximate Bayes' factors were then summed across the 1-megabase (Mb) region and the proportion of the posterior odds of being the causal variant was calculated for each variant (approximate Bayes' factor for SNP/sum of approximate Bayes' factors for the region). The set of SNPs that accounts for 99% of posterior odds of association in the region

denotes the set most likely to contain the causal variant for that association region (Supplementary Table 12).

Cumulative effects, risk prediction and variance explained. We assessed the cumulative effects of the 97 GWS loci on mean BMI and on their ability to predict obesity ($\text{BMI} \geq 30 \text{ kg m}^{-2}$) using the c statistic from logistic regression models in the Health and Retirement Study¹⁷, a longitudinal study of 26,000 European Americans 50 years or older. The variance explained (VarExp) by each SNP was calculated using the effect allele frequency (f) and beta (β) from the meta-analyses using the formula $\text{VarExp} = \beta^2(1-f)2f$.

For polygene analyses, the approximate conditional analysis from GCTA^{19,20}, was used to select SNPs using a range of P value thresholds (that is, 5×10^{-8} , 5×10^{-7} , ..., 5×10^{-3}) based on summary data from the European sex-combined meta-analysis excluding TwinGene and QIMR studies. We performed a within-family prediction analysis using full-sib pairs selected from independent families (1,622 pairs from the QIMR cohort and 2,758 pairs from the TwinGene cohort) and then SNPs at each threshold were used to calculate the percentage of phenotypic variance explained and predict risk (Extended Data Figs 2 and 3). We then confirmed the results from population-based prediction and estimation analyses in an independent sample of unrelated individuals from the TwinGene ($n = 5,668$) and QIMR ($n = 3,953$) studies (Extended Data Fig. 3 and Fig. 1c). The SNP-derived predictor was calculated using the profile scoring approach implemented in PLINK and estimation analyses were performed using the all-SNP estimation approach implemented in GCTA.

Enrichment analysis of Metabochip SNPs selected for replication. The 5,055 SNPs that were included for BMI replication on Metabochip included 1,909 independent SNPs ($r^2 < 0.1$ and > 500 kb apart), of which 1,458 displayed directionally consistent effect estimates with those reported previously⁵. To estimate the number of Metabochip SNPs truly associated with BMI, we counted the number of SNPs with directional consistency (DC) between ref. 5 and a meta-analysis of non-overlapping samples for these 1,909 SNPs. We then calculated DC in the presence of a mixture of associated and non-associated SNPs assuming $P(\text{DC} | \text{associated}) = 1$ and $P(\text{DC} | \text{not associated}) = 0.5$. In this formulation, $\text{DC} = R/2 + S$, meaning that $S = 2\text{DC} - T$, in which T equals the total number of SNPs, R equals the number of SNPs not associated with BMI, and S equals the number of SNPs associated with BMI. With $\text{DC} = 1,458$ and $T = 1,909$, we estimate S to be $2\text{DC} - T = 2 \times 1,458 - 1,909 = 1,007$.

Joint and conditional multiple SNP association analysis. To identify additional signals in regions of association, we used GCTA¹⁹, an approach that uses meta-analysis summary statistics and an LD matrix derived from a reference sample, to perform approximate joint and conditional SNP association analysis. We used 6,654 unrelated individuals of European ancestry from the ARIC cohort as the reference sample to approximate conditional P values.

Manual gene annotation and biological description. All genes within 500 kb of an index SNP were annotated for molecular function, cellular function, and for evidence of association with BMI-related traits in human or animal model experiments (Supplementary Table 22). We used several avenues for annotation, including Spotter (<http://csg.sph.umich.edu/boehnke/spotter/>), SNIPPER (<http://csg.sph.umich.edu/boehnke/snipper/>), PubMed (<http://www.ncbi.nlm.nih.gov/pubmed/>), OMIM (<http://www.omim.org>) and UNIPROT (<http://www.uniprot.org/>). When no genes mapped to this interval the nearest gene on each side of the index SNP was annotated. In examining possible functions of genes in the region, we excluded any references to GWAS or other genetic association studies. We analysed 405 genes in the 97 GWS loci and manually curated them into 25 biological categories containing more than three genes.

Functional variants. All variants within 500 kb (HapMap release 22/1000 Genomes CEU) and in LD ($r^2 > 0.7$) with an index SNP were annotated for functional effects based on RefSeq transcripts using Annovar⁴³ (<http://www.openbioinformatics.org/annovar/>). PhastCon, Grantham, GERP, and PolyPhen⁴⁴ predictions were accessed via the Exome Variant Server⁴⁵ (<http://evs.gs.washington.edu/EVS/>), and from SIFT⁴⁶ (<http://sift.jcvi.org/>) (Extended Data Table 4).

Copy number variations correlated with BMI index SNPs. To study common copy number variations, we used a list of copy number variations well-tagged by SNPs in high LD ($r^2 > 0.8$) with deletions in European populations from phase 1 release of the 1000 Genomes Project⁴⁷ (Supplementary Table 25).

eQTLs. We examined the *cis* associations between the 97 GWS SNPs and expression of nearby genes in whole blood, lymphocytes, skin, liver, omental fat, subcutaneous fat and brain tissue^{48–55} (Supplementary Table 23). Conditional analyses were performed by including both the BMI-associated SNP and the most significant *cis*-associated SNP for the given transcript. Conditional analyses were conducted for all data sets, except the brain tissue data set due to limited power. To minimize the potential for false-positives, only *cis* associations below a study-specific FDR of 5% (or 1% for some data sets), in LD with the peak SNP ($r^2 > 0.7$) for the transcript, and with conditional $P > 0.05$ for the peak SNP, are reported (Extended Data Table 2).

MAGENTA. We used the MAGENTA method to test predefined gene sets for enrichment at BMI-associated loci²⁴. We used the GWAS + Metabochip data as input and applied default settings.

GRAIL. We used GRAIL²⁷ to identify genes near BMI-associated loci having similarities in the published scientific text using PubMed abstracts as of December 2006. The BMI loci were queried against HapMap release 22 for the European panel, and we controlled for gene size.

DEPICT. We used DEPICT to identify the most likely causal gene at a given associated locus, reconstituted gene sets enriched for BMI associations, and tissues and cell types in which genes from associated loci are highly expressed²¹. To accomplish this, the method relies on publicly available gene sets (including molecular pathways) and uses gene expression data from 77,840 gene expression arrays⁷⁵ to predict which other genes are likely to be part of these gene sets, thus combining known annotations with predicted annotations. For details and negative control analyses please see Supplementary Methods.

We first clumped the European-only GWAS-based meta-analysis summary statistics using 500 kb flanking regions, LD $r^2 > 0.1$ and excluded SNPs with $P \geq 5 \times 10^{-4}$; which resulted in a list of 590 independent SNPs. HapMap phase II CEU genotype data³⁷ was used to compute LD and genomic coordinates were defined by genome build GRCh38. Because the GWAS meta-analysis was based on both GWAS and Metabochip studies, there were discrepancies in the index SNPs that are referenced in Table 1 of the paper and the ones used in DEPICT, which was run on the GWAS data only. Therefore we forced in GWS index SNPs from the GWAS plus Metabochip GWA meta-analysis into the DEPICT GWAS-only based analysis. This enabled a more straightforward comparison of genes in DEPICT loci and genes in GWS loci highlighted by manual lookups, and did not lead to any significant bias towards SNPs on Metabochip (data not shown). We forced in 62 of the GWS loci in Table 1, so all of the 97 SNPs were among the 590 SNPs. The 590 SNPs were further merged into 511 non-overlapping regions (FDR < 0.05) used in DEPICT analysis. For additional information on the analysis please refer to Supplementary Methods.

Cross-trait analyses. To explore the relationship between BMI and an array of cardiometabolic traits and diseases, association results for the 97 BMI index SNPs were requested from 13 GWAS meta-analysis consortia: DIAGRAM (type 2 diabetes)⁵⁶, CARDIOGRAM-C4D (CAD)⁵⁷, ICBP (systolic and diastolic blood pressure (SBP, DBP))⁵⁸, GIANT (waist-to-hip ratio, hip circumference, and waist circumference, each unadjusted and adjusted for BMI)^{13,39}, GLGC (HDL, low density lipoprotein cholesterol, triglycerides, and total cholesterol)⁶⁰, MAGIC (fasting glucose, fasting insulin, fasting insulin adjusted for BMI, and two-hour glucose)^{61–63}, ADIPOGen (BMI-adjusted adiponectin)⁶⁴, CKDgen (urine albumin-to-creatinine ratio (UACR), estimated glomerular filtration rate, and overall CKD)^{65,66}, ReproGen (age at menarche, age at menopause)^{67,68}, GENIE (diabetic nephropathy)^{69,70}. Proxies ($r^2 > 0.8$ in CEU) were used when an index SNP was unavailable.

Enrichment of concordant effects. We compared the effects for the 97 BMI index SNP across these related traits using a one-sided binomial test of the number of concordant effects versus a null expectation of $P = 0.5$. Concordant and nominally significant ($P < 0.05$) SNP effects were similarly tested using a one-sided binomial test with a null expectation of $P = 0.05$. We evaluated significance in either test with a Bonferroni-corrected threshold of $P = 0.002$ (0.05/23 traits tested).

Joint effects of cross-trait associations. To determine the joint effect of all 97 BMI loci on other cardiometabolic phenotypes, we used the meta-regression technique from ref. 64 to correlate the effect estimates of the BMI-increasing alleles with effect estimates from meta-analyses for each of the metabolic traits from other consortia (DIAGRAM, MAGIC, ICBP, GLGC, ADIPOGen, ReproGen and CARDIOGRAM).

Cross-traits heatmap. To explore observed concordance in effects of BMI loci on other cardiometabolic and anthropometric traits, we converted the effect estimates and standard errors (or P values) from meta-analysis to Z -scores oriented with respect to the BMI-increasing allele, for each of the 97 BMI index SNPs in the twenty-three traits. We then classified each Z -score as follows to generate a vector of the Z -score of each trait at each locus:

0 (not significant) if $-2 \leq Z \leq 2$;

1 (significant positive) if $Z > 2$;

–1 (significant negative) if $Z < -2$.

Extended Data Fig. 5 displays these locus-trait relationships in a heatmap using Euclidean distance and complete linkage clustering to order both loci and traits.

Cross-traits bubble plot. We also represent the genetic overlap between other cardiometabolic traits and BMI susceptibility loci with a bubble plot in which the size of each bubble is proportional to the fraction of BMI-associated loci for which there was a significant association ($P < 5 \times 10^{-4}$). Each pair of bubbles is connected by a line proportional to the number of significant BMI-increasing loci overlapping between the traits.

NHGRI GWAS catalogue lookups. We extracted previously reported GWAS association within 500 kb of and $r^2 > 0.7$ with any BMI-index SNP from the NHGRI GWAS catalogue⁷¹ (<http://www.genome.gov/gwastudies>; Supplementary Table 17a,

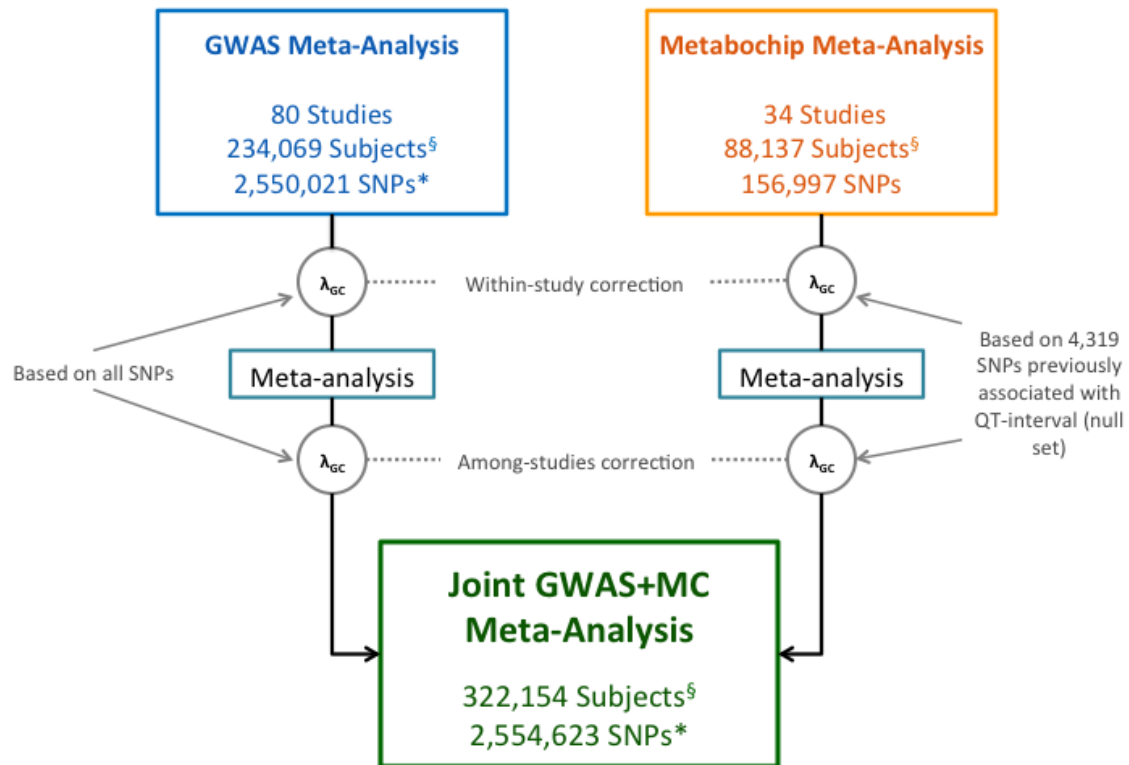
b). For studies reporting greater than 30 significant hits, additional SNP-trait associations were pulled from the literature and compared to BMI index SNPs the same as with other GWAS catalogue studies.

ENCODE/Roadmap. To identify global enrichment of data sets at the BMI-associated loci we performed permutation-based tests in a subset of 41 open chromatin (DNase-seq), histone modification (H3K27ac, H3K4me1, H3K4me3 and H3K9ac), and transcription factor binding data sets from the ENCODE Consortium²², Roadmap Epigenomics Project²³ and when available the ENCODE Integrative Analysis^{60,72} (Supplementary Table 19). We processed Roadmap Epigenomics sequencing data with multiple biological replicates using MACS2 (ref. 73) and then applied same Irreproducible Discovery Rate pipeline used in the ENCODE Integrative Analysis^{60,72}. Roadmap Epigenomics data with only a single replicate were analysed using MACS2 alone. We examined variants in LD with 97 BMI index SNPs based on $r^2 > 0.7$ from the 1000 Genomes phase 1 version 2 EUR samples⁷⁴. We matched the index SNP at each locus with 500 variants having no evidence of association ($P > 0.5$, ~1.2 million total variants) with a similar distance to the nearest gene ($\pm 11,655$ bp), number of variants in LD (± 8 variants), and minor allele frequency. Using these pools, we created 10,000 sets of control variants for each of the 97 loci and identified variants in LD ($r^2 > 0.7$) and within 1 Mb. For each SNP set, we calculated the number of loci with at least one variant located in a regulatory region under the assumption that one regulatory variant is responsible for each association signal. We estimated the P value assuming a sum of binomial distributions to represent the number of index SNPs (or their LD proxies; $r^2 > 0.7$) that overlap a regulatory data set compared to the expectation observed in the 500 matched control sets. Data sets were considered significantly enriched if the P value was below a Bonferroni-corrected threshold of 1.2×10^{-3} , adjusting for 41 tests.

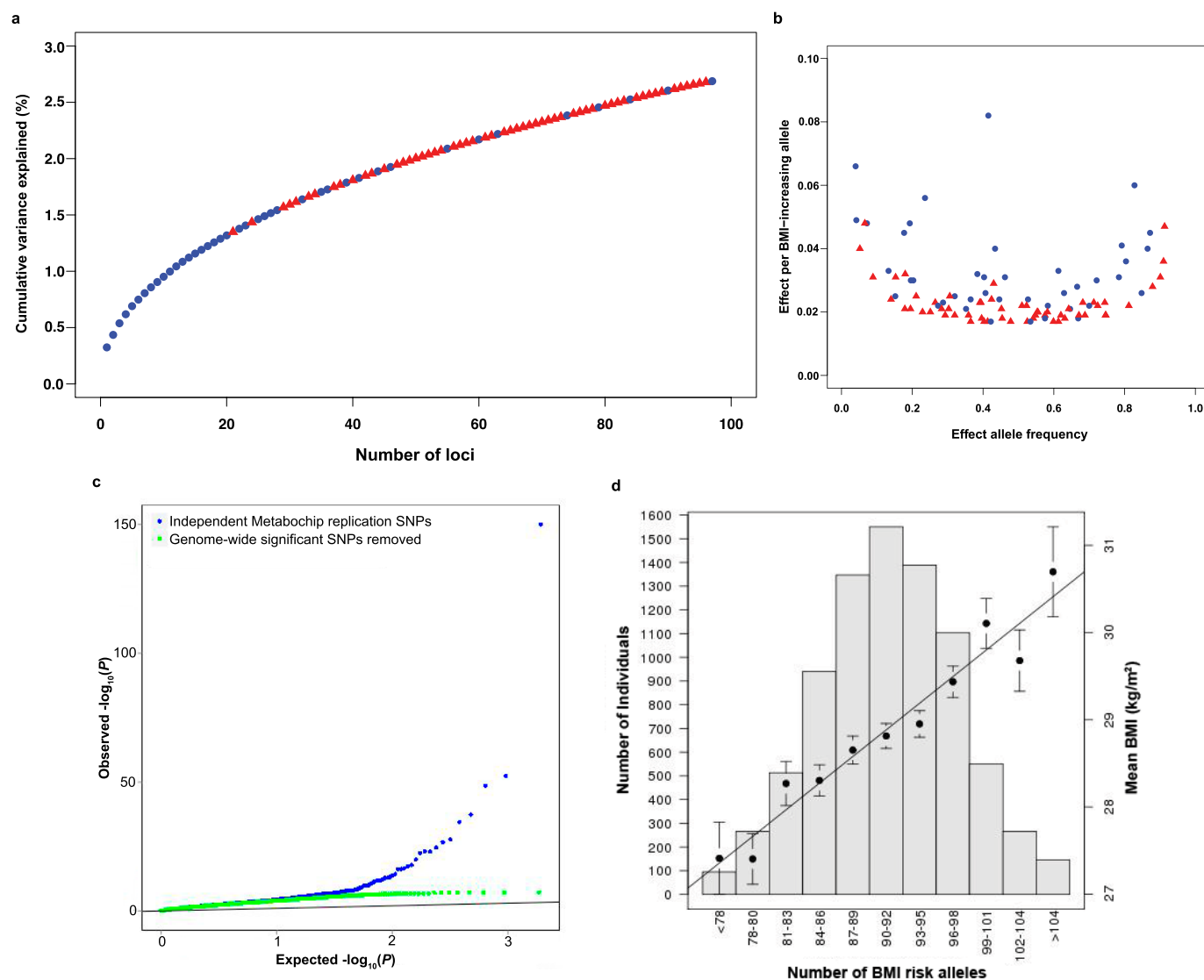
37. Frazer, K. A. *et al.* A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449**, 851–861 (2007).
38. Winkler, T. W. *et al.* Quality control and conduct of genome-wide association meta-analyses. *Nature Protocols* **9**, 1192–1212 (2014).
39. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
40. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55**, 997–1004 (1999).
41. Wen, W. *et al.* Meta-analysis identifies common variants associated with body mass index in east Asians. *Nature Genet.* **44**, 307–311 (2012).
42. Randall, J. C. *et al.* Sex-stratified genome-wide association studies including 270,000 individuals show sexual dimorphism in genetic loci for anthropometric traits. *PLoS Genet.* **9**, e1003500 (2013).
43. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164 (2010).
44. Adzhubei, I. A. *et al.* A method and server for predicting damaging missense mutations. *Nature Methods* **7**, 248–249 (2010).
45. NHLBI Exome Sequencing Project (ESP), Exome Variant Server; <http://evs.gs.washington.edu/EVS/>.
46. Ng, P. C. & Henikoff, S. Predicting deleterious amino acid substitutions. *Genome Res.* **11**, 863–874 (2001).
47. Mills, R. E. *et al.* Mapping copy number variation by population-scale genome sequencing. *Nature* **470**, 59–65 (2011).
48. Emilsson, V. *et al.* Genetics of gene expression and its effect on disease. *Nature* **452**, 423–428 (2008).
49. Zhong, H., Yang, X., Kaplan, L. M., Molony, C. & Schadt, E. E. Integrating pathway analysis and genetics of gene expression for genome-wide association studies. *Am. J. Hum. Genet.* **86**, 581–591 (2010).
50. Grundberg, E. *et al.* Mapping cis- and trans-regulatory effects across multiple tissues in twins. *Nature Genet.* **44**, 1084–1089 (2012).
51. Dixon, A. L. *et al.* A genome-wide association study of global gene expression. *Nature Genet.* **39**, 1202–1207 (2007).
52. Fehrmann, R. S. *et al.* Trans-eQTLs reveal that independent genetic variants associated with a complex phenotype converge on intermediate genes, with a major role for the HLA. *PLoS Genet.* **7**, e1002197 (2011).
53. Nelis, M. *et al.* Genetic structure of Europeans: a view from the North-East. *PLoS ONE* **4**, e5472 (2009).
54. Myers, A. J. *et al.* A survey of genetic human cortical gene expression. *Nature Genet.* **39**, 1494–1499 (2007).
55. Westra, H. J. *et al.* Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nature Genet.* **45**, 1238–1243 (2013).
56. Morris, A. P. *et al.* Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nature Genet.* **44**, 981–990 (2012).
57. Deloukas, P. *et al.* Large-scale association analysis identifies new risk loci for coronary artery disease. *Nature Genet.* **45**, 25–33 (2013).
58. Ehret, G. B. *et al.* Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature* **478**, 103–109 (2011).
59. Shungin, D. *et al.* New genetic loci link adipose and insulin biology to body fat distribution. *Nature* <http://dx.doi.org/nature14132> (this issue).
60. Willer, C. *et al.* Discovery and refinement of loci associated with lipid levels. *Nature Genet.* **45**, 1274–1283 (2013).

61. Scott, R. A. *et al.* Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. *Nature Genet.* **44**, 991–1005 (2012).
62. Manning, A. K. *et al.* A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. *Nature Genet.* **44**, 659–669 (2012).
63. Saxena, R. *et al.* Genetic variation in GIPR influences the glucose and insulin responses to an oral glucose challenge. *Nature Genet.* **42**, 142–148 (2010).
64. Dastani, Z. *et al.* Novel loci for adiponectin levels and their influence on type 2 diabetes and metabolic traits: a multi-ethnic meta-analysis of 45,891 individuals. *PLoS Genet.* **8**, e1002607 (2012).
65. Pattaro, C. *et al.* Genome-wide association and functional follow-up reveals new loci for kidney function. *PLoS Genet.* **8**, e1002584 (2012).
66. Böger, C. A. *et al.* CUBN is a gene locus for albuminuria. *J. Am. Soc. Nephrol.* **22**, 555–570 (2011).
67. Stolk, L. *et al.* Meta-analyses identify 13 loci associated with age at menopause and highlight DNA repair and immune pathways. *Nature Genet.* **44**, 260–268 (2012).
68. Elks, C. E. *et al.* Thirty new loci for age at menarche identified by a meta-analysis of genome-wide association studies. *Nature Genet.* **42**, 1077–1085 (2010).
69. Williams, W. W. *et al.* Association testing of previously reported variants in a large case-control meta-analysis of diabetic nephropathy. *Diabetes* **61**, 2187–2194 (2012).
70. Sandholm, N. *et al.* New susceptibility loci associated with kidney disease in type 1 diabetes. *PLoS Genet.* **8**, e1002921 (2012).
71. Hindorf, L. A. *et al.* Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl Acad. Sci. USA* **106**, 9362–9367 (2009).
72. Li, Q., Brown, J. B., Huang, H. & Bickel, P. J. Measuring reproducibility of high-throughput experiments. *Ann. Appl. Stat.* **5**, 1752–1779 (2011).
73. Feng, J., Liu, T., Qin, B., Zhang, Y. & Liu, X. S. Identifying ChIP-seq enrichment using MACS. *Nature Protocols* **7**, 1728–1740 (2012).
74. Abecasis, G. R. *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012).
75. Fehrmann, R. S. *et al.* Gene expression analysis identifies global gene dosage sensitivity in cancer. *Nature Genet.* **47**, 115–125 (2015).

European Studies Only

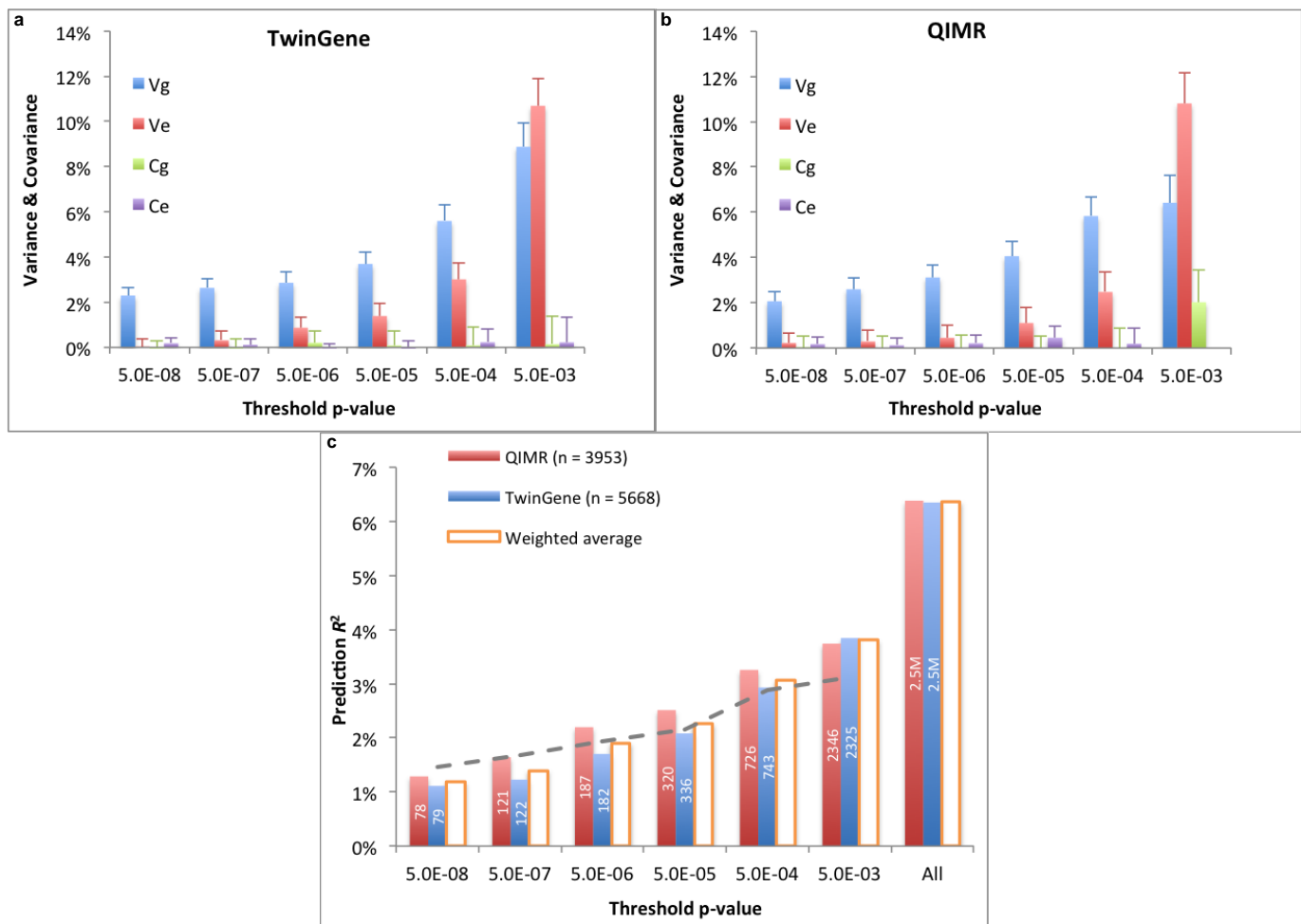


Extended Data Figure 1 | Study design. *The SNP counts reflect sample size filter of $n \geq 50,000$. [§]Counts represent the primary European sex-combined analysis. Please see Extended Data Table 1 for counts for secondary analyses.



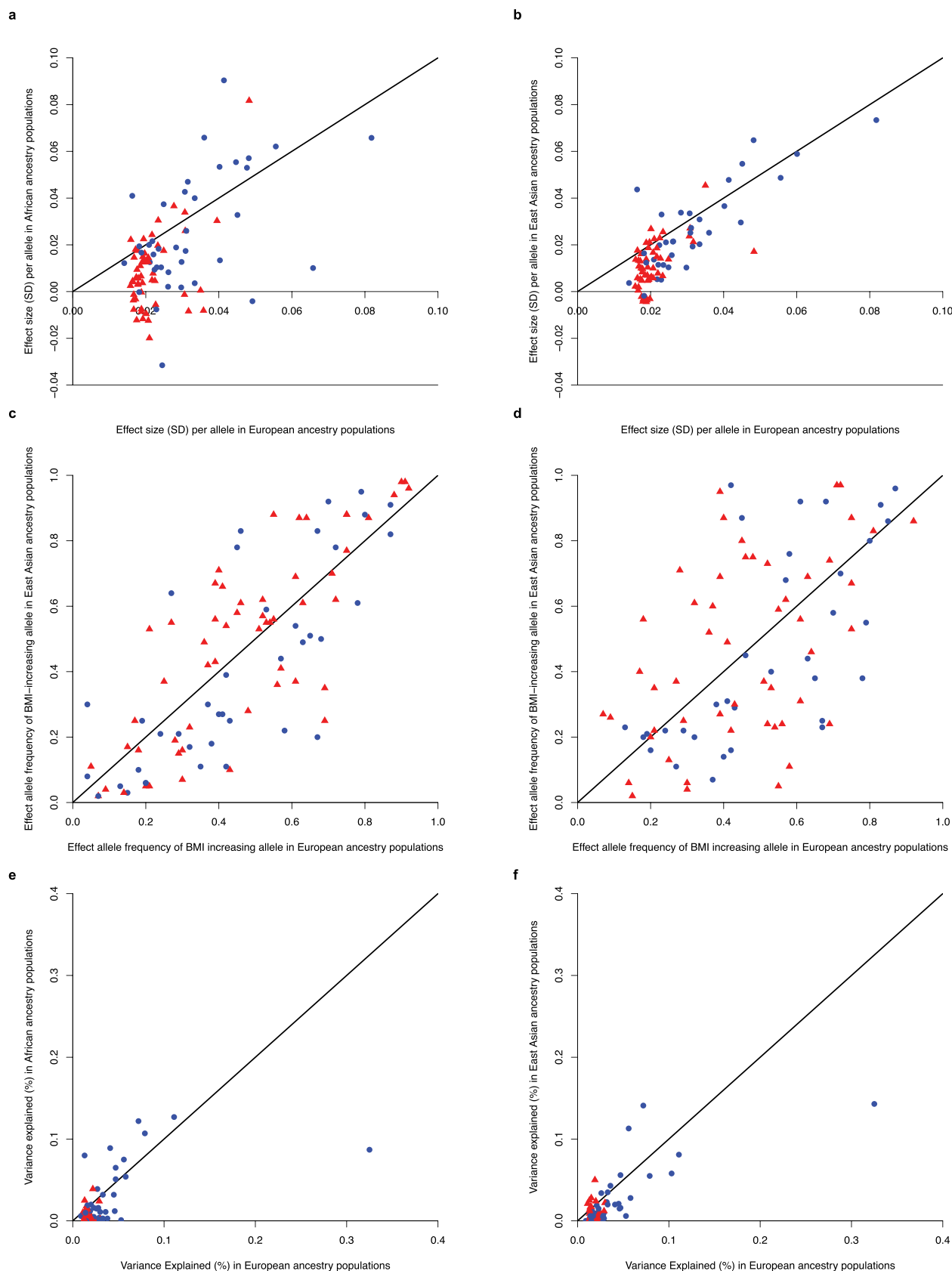
Extended Data Figure 2 | Genetic characterization of BMI-associated variants. **a**, Plot of the cumulative phenotypic variance explained by each locus ordered by decreasing effect size. **b**, The relationship between effect size and allele frequency. Previously identified loci are blue circles and novel loci are red triangles. **c**, Quantile–quantile (Q–Q) plot of meta-analysis P values for all

1,909 BMI-replication SNPs (blue) and after removing SNPs near the 97 associated loci (green). **d**, Histogram of cumulative effect of BMI risk alleles. Mean BMI for each bin is shown by the black dots (with standard deviation) and corresponds to the right-hand y-axis.



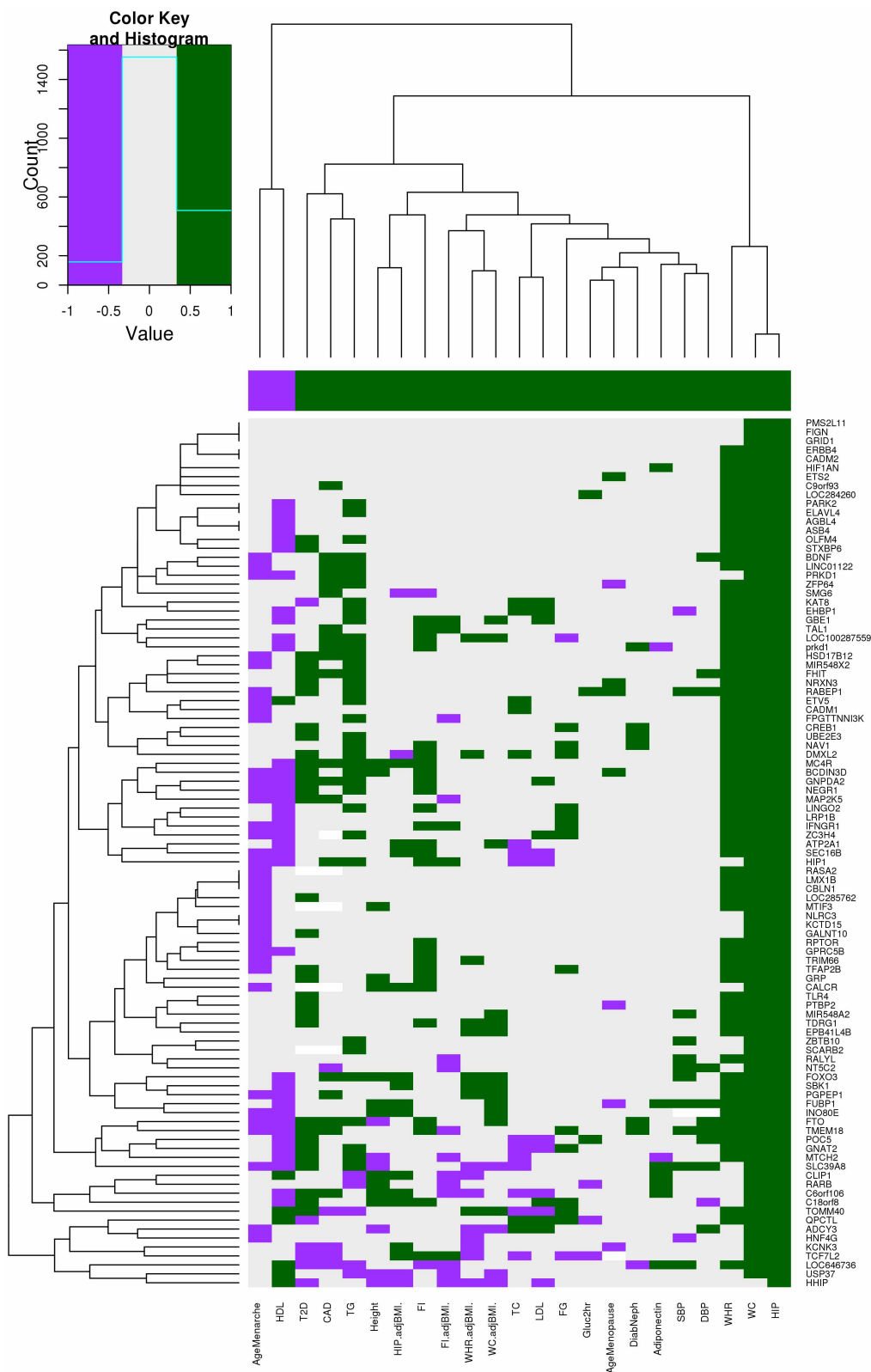
Extended Data Figure 3 | Partitioning the variance in and risk prediction from SNP-derived predictor. **a, b,** The analyses were performed using 2,758 full sibling pairs from the TwinGene cohort (**a**) and 1,622 pairs from the QIMR cohort (**b**). The SNP-based predictor was adjusted for the first 20 principal components. The variance of the SNP-based predictor can be partitioned into four components (V_g , V_e , C_g and C_e) using the within-family prediction analysis, in which V_g is the variance explained by real SNP effects, C_g is the covariance between predictors attributed to the real effects of SNPs that are not in LD but correlated due to population stratification, V_e is the accumulated variance due to the errors in estimating SNP effects, and C_e is the covariance between predictors attributed to errors in estimating the effects of SNPs that are

correlated due to population stratification. Error bars reflect s.e.m. of estimates. **c,** The prediction R^2 shown on the y axis is the squared correlation between phenotype and SNP-based genetic predictor in unrelated individuals from the TwinGene ($n = 5,668$) and QIMR ($n = 3,953$) studies. The number shown in each column is the number of SNPs selected from the GCTA joint and conditional analysis at a range of P -value thresholds. In each case, the predictor was adjusted by the first 20 principal components. The column in orange is the average prediction R^2 weighted by sample size over the two cohorts. The dashed grey line is the value inferred from the within-family prediction analyses using this equation $R^2 = (V_g + C_g)^2 / (V_g + V_e + C_g + C_e)$.



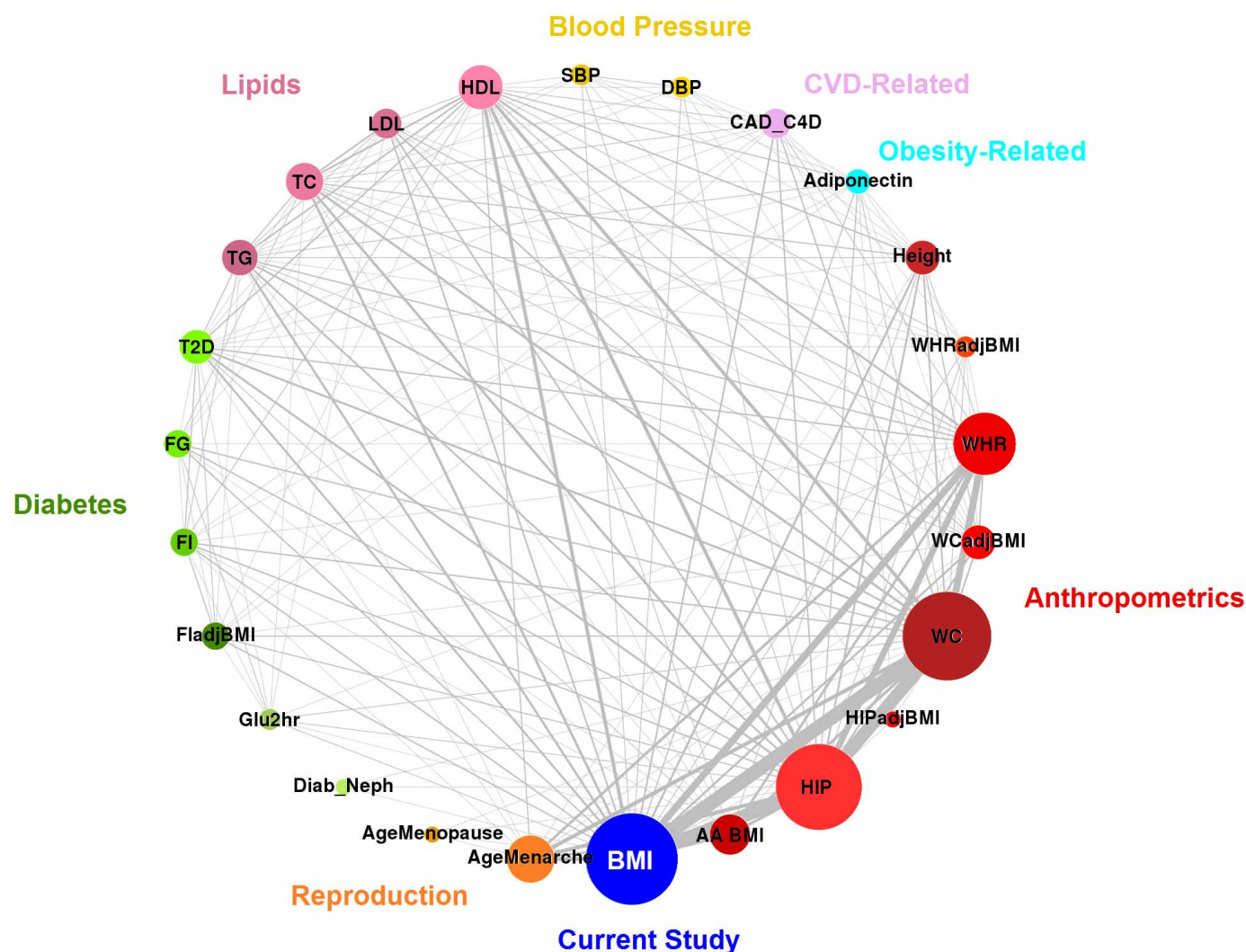
Extended Data Figure 4 | Comparison of BMI-associated index SNPs across ethnicities. **a, b,** BMI effects observed in European ancestry individuals (x axes) compared to African ancestry (**a**) or Asian ancestry (**b**) individuals

(y axes). **c, d,** Allele frequencies between ancestry groups, as in **a** and **b**. **e, f,** Comparison of the estimates of explained variance. In all plots, novel loci are in red and previously identified loci are in blue.



Extended Data Figure 5 | Effects of BMI-associated loci on related metabolic traits. Unsupervised hierarchical clustering of the 97 BMI-associated loci (y axis) on 23 related metabolic traits (x axis). The top row shows the a priori expected relationship with BMI (green is concordant effect direction, purple is opposite). Loci with statistically significant concordant

direction of effect are highlighted in green, and significant but opposing effects are in purple. Grey indicates a non-significant relationship and those with no information are in white. The key in the top left corner also shows the count of gene-phenotype pairs in each category (cyan bars).



Extended Data Figure 6 | Bubble chart representing the genetic overlap across traits at BMI susceptibility loci. Each bubble represents a trait for which association results were requested for the 97 GWS BMI loci. The size of the bubble is proportional to the number of BMI-increasing loci with a significant association. A line connects each pair of bubbles with thickness proportional to the number of significant loci shared between the traits. Traits tested include the current study BMI SNPs, African-American BMI (AA BMI), hip circumference (HIP), HIP adjusted for BMI (HIPadjBMI), waist circumference (WC), waist circumference adjusted for BMI (WCadjBMI),

waist-to-hip ratio (WHR), waist-to-hip ratio adjusted for BMI (WHRadjBMI), height, adiponectin, coronary artery disease (CAD), diastolic blood pressure (DBP), systolic blood pressure (SBP), high-density lipoprotein (HDL), low-density lipoprotein (LDL), total cholesterol (TC), triglycerides (TG), type 2 diabetes (T2D), fasting glucose (FG), fasting insulin (FI), fasting insulin adjusted for BMI (FIadjBMI), two-hour glucose (Glu2hr), diabetic nephropathy (Diab_Neph), age at menopause (AgeMenopause), and age at menarche (AgeMenarche).

Extended Data Table 1 | Descriptive characteristics of meta-analyses

Meta-analysis	Total number of studies	Maximum number of subjects	Number of SNPs*	λ_{GC}
<i>European sex-combined</i>				
GWAS	80	234,069	2,550,021	1.526
Metabochip	34	88,137	156,997	1.25
Joint GWAS+Metabochip	114	322,154	2,554,623	1.084
<i>European men</i>				
GWAS	72	104,666	2,473,152	1.279
Metabochip	34	48,274	152,326	1.121
Joint GWAS+Metabochip	106	152,893	2,477,617	1.006
<i>European women</i>				
GWAS	74	132,115	2,491,697	1.336
Metabochip	33	39,864	153,086	1.029
Joint GWAS+Metabochip	107	171,977	2,494,571	1.002
<i>European population-based</i>				
GWAS	49	162,262	2,502,573	1.385
Metabochip	20	46,263	155,617	1.034
Joint GWAS+Metabochip	69	209,521	2,506,448	1.003
<i>All ancestries</i>				
GWAS	82	236,231	2,550,614	1.451
Metabochip	43	103,047	181,718	1.25
Joint GWAS+Metabochip	125	339,224	2,555,496	1.004

* For the GWAS and joint GWAS+Metabochip analyses, SNP count reflects $n \geq 50,000$.

Extended Data Table 2 | Previously known GWS BMI loci in European meta-analysis

SNP	Chr:Position	*Notable gene(s)	Alleles	EAF	β	SE	P value
rs1558902	16:52,361,075	<i>FTO</i> (B,N)	A/T	0.415	0.082	0.003	7.51E-153
rs6567160	18:55,980,115	<i>MC4R</i> (B,N)	C/T	0.236	0.056	0.004	3.93E-53
rs13021737	2:622,348	<i>TMEM18</i> (N)	G/A	0.828	0.06	0.004	1.11E-50
rs10938397	4:44,877,284	<i>GNPDA2</i> (N); <i>GABRG1</i> (B)	G/A	0.434	0.04	0.003	3.21E-38
rs543874	1:176,156,103	<i>SEC16B</i> (N)	G/A	0.193	0.048	0.004	2.62E-35
rs2207139	6:50,953,449	<i>TFAP2B</i> (B,N)	G/A	0.177	0.045	0.004	4.13E-29
rs11030104	11:27,641,093	<i>BDNF</i> (B,M,N)	A/G	0.792	0.041	0.004	5.56E-28
rs3101336	1:72,523,773	<i>NEGR1</i> (B,C,D,N)	C/T	0.613	0.033	0.003	2.66E-26
rs7138803	12:48,533,735	<i>BCDIN3D</i> (N); <i>FAIM2</i> (D)	A/G	0.384	0.032	0.003	8.15E-24
rs10182181	2:25,003,800	<i>ADCY3</i> (B,M,N,Q); <i>POMC</i> (B,G); <i>NCOA1</i> (B)	G/A	0.462	0.031	0.003	8.78E-24
rs3888190	16:28,796,987	<i>SH2B1</i> (B,M,Q); <i>APOBR</i> (M,Q); <i>ATXN2L</i> (Q); <i>SBK1</i> (Q,D); <i>SULT1A2</i> (Q); <i>TUFM</i> (Q)	A/C	0.403	0.031	0.003	3.14E-23
rs1516725	3:187,306,698	<i>ETV5</i> (N)	C/T	0.872	0.045	0.005	1.89E-22
rs12446632	16:19,842,890	<i>GPRC5B</i> (C,N); <i>IQCK</i> (Q)	G/A	0.865	0.04	0.005	1.48E-18
rs2287019	19:50,894,012	<i>QPCTL</i> (N); <i>GIPR</i> (B,M)	C/T	0.804	0.036	0.004	4.59E-18
rs16951275	15:65,864,222	<i>MAP2K5</i> (B,D,N); <i>LBXCOR1</i> (M)	T/C	0.784	0.031	0.004	1.91E-17
rs3817334	11:47,607,569	<i>MTCH2</i> (M,Q); <i>C1QTNF4</i> (Q,I); <i>SPI1</i> (Q); <i>CELF1</i> (D)	T/C	0.407	0.026	0.003	5.15E-17
rs2112347	5:75,050,998	<i>POC5</i> (M); <i>HMGCR</i> (B); <i>COL4A3BP</i> (B)	T/G	0.629	0.026	0.003	6.19E-17
rs12566985	1:74,774,781	<i>FPGT-TNNI3K</i> (N)	G/A	0.446	0.024	0.003	3.28E-15
rs3810291	19:52,260,843	<i>ZC3H4</i> (D,N,Q)	A/G	0.666	0.028	0.004	4.81E-15
rs7141420	14:78,969,207	<i>NRXN3</i> (D,N)	T/C	0.527	0.024	0.003	1.23E-14
rs13078960	3:85,890,280	<i>CADM2</i> (D,N)	G/T	0.196	0.03	0.004	1.74E-14
rs10968576	9:28,404,339	<i>LINGO2</i> (D,N)	G/A	0.32	0.025	0.003	6.61E-14
rs17024393	1:109,956,211	<i>GNAT2</i> (N); <i>AMPD2</i> (D)	C/T	0.04	0.066	0.009	7.03E-14
rs12429545	13:53,000,207	<i>OLFM4</i> (B,N)	A/G	0.133	0.033	0.005	1.09E-12
rs13107325	4:103,407,732	<i>SLC39A8</i> (M,N,Q)	T/C	0.072	0.048	0.007	1.83E-12
rs11165643	1:96,696,685	<i>PTBP2</i> (D,N)	T/C	0.583	0.022	0.003	2.07E-12
rs17405819	8:76,969,139	<i>HNF4G</i> (B,N)	T/C	0.7	0.022	0.003	2.07E-11
rs1016287	2:59,159,129	<i>LINC01122</i> (N)	T/C	0.287	0.023	0.003	2.25E-11
rs4256980	11:8,630,515	<i>TRIM66</i> (D,M,N); <i>TUB</i> (B)	G/C	0.646	0.021	0.003	2.90E-11
rs12401738	1:78,219,349	<i>FUBP1</i> (N); <i>USP33</i> (D)	A/G	0.352	0.021	0.003	1.15E-10
rs205262	6:34,671,142	<i>C6orf106</i> (N); <i>SNRPC</i> (Q)	G/A	0.273	0.022	0.004	1.75E-10
rs12016871	13:26,915,782	<i>MTIF3</i> (N); <i>GTF3A</i> (Q)	T/C	0.203	0.03	0.005	2.29E-10
rs12940622	17:76,230,166	<i>RPTOR</i> (B,N)	G/A	0.575	0.018	0.003	2.49E-09
rs11847697	14:29,584,863	<i>PRKD1</i> (N)	T/C	0.042	0.049	0.008	3.99E-09
rs2075650	19:50,087,459	<i>TOMM40</i> (B,N); <i>APOE</i> (B); <i>APOC1</i> (B)	A/G	0.848	0.026	0.005	1.25E-08
rs2121279	2:142,759,755	<i>LRP1B</i> (N)	T/C	0.152	0.025	0.004	2.31E-08
rs29941	19:39,001,372	<i>KCTD15</i> (N)	G/A	0.669	0.018	0.003	2.41E-08
rs1808579	18:19,358,886	<i>NPC1</i> (B,G,M,Q); <i>C18orf8</i> (N,Q)	C/T	0.534	0.017	0.003	4.17E-08

SNP positions are reported according to Build 36 and their alleles are coded based on the positive strand. Effect alleles, allele frequencies, betas (β), s.e.m., sample sizes (n), and P values are based on the meta-analysis of GWAS I+II+MetaboChIP association data from the European sex-combined data set.

* Notable genes from biological relevance to obesity (B); GRAIL results (G); BMI-associated variant is in strong LD ($r^2 \geq 0.7$) with a missense variant in the indicated gene (M); gene nearest to Index SNP (N); association and eQTL data converge to affect gene expression (Q); DEPICT analyses (D); copy number variation (C).

Extended Data Table 3 | Association of the GWS SNPs for BMI with *cis*-gene expression (*cis*-eQTLs)

SNP	Chr.	BMI	Tissue	Gene	β for	P for	P_{adj} for	Peak SNP	r^2	P for peak	P_{adj} for	Reference
		increasing allele			Giant SNP		GIANT SNP			SNP	SNP	
Novel loci												
rs11583200	1	C	Subcutaneous	ELAVL4	-0.066	1.90E-12	0.44	rs6588374	0.78	1.07E-12	0.36	Zhong et al.
rs492400	2	C	Liver	PLCD4	-0.054	4.64E-40	0.98	rs10187066	1.00	4.49E-40	0.98	Zhong et al.
rs492400	2	C	Lymphocyte	RQCD1	0.392	7.11E-22	0.94	rs526134	1.00	4.06E-22	0.21	Dixon et al.
rs492400	2	C	PBMC	RQCD1	-0.102	2.43E-06	0.98	rs526134	0.95	2.21E-06	0.96	PBMC meta-analysis
rs492400	2	C	Omental	TTL4	0.018	1.33E-10	0.82	rs12987009	0.73	2.82E-13	0.07	
rs492400	2	C	Lymphocyte	TTL4	0.158	9.02E-06	1	rs492400	1.00	9.02E-06	1	Dixon et al.
rs17001654	4	G	Lymphocyte	SCARB2	0.248	5.57E-09	0.59	rs6835324	0.94	3.42E-09	0.25	Dixon et al.
rs9400239	6	C	Subcutaneous	HSS00296402	0.034	9.51E-22	0.97	rs2153960	0.94	1.93E-23	0.48	Zhong et al.
rs9400239	6	C	Omental	HSS00296402	0.015	1.34E-13	0.50	rs2153960	0.93	4.64E-17	0.22	Zhong et al.
rs1167827	7	G	Blood	PMS2P3	-0.595	4.20E-32	0.66	rs6963105	0.93	3.00E-32	0.39	Emilsson et al.
rs1167827	7	G	Omental	PMS2P3	-0.027	1.57E-11	0.95	rs6963105	0.98	6.94E-12	0.86	Zhong et al.
rs1167827	7	G	Subcutaneous	PMS2P3	-0.030	1.30E-10	0.71	rs1167796	0.73	1.04E-12	0.10	Zhong et al.
rs1167827	7	G	Adipose	PMS2P3	-0.346	3.40E-09	1	rs1167827	1.00	3.40E-09	1	Emilsson et al.
rs1167827	7	G	Blood	PMS2P5	-0.367	1.20E-11	0.47	rs6963105	0.93	5.00E-12	0.14	Emilsson et al.
rs1167827	7	G	Subcutaneous	WBSCR16	0.025	1.44E-10	1	rs1167827	1.00	1.44E-10	1	Zhong et al.
rs1167827	7	G	Omental	WBSCR16	0.017	1.75E-06	1	rs1167827	1.00	1.75E-06	1	Zhong et al.
rs9641123	7	C	Abdominal SAT	hsa-miR-653	-0.344	1.54E-04	0.23	rs16868443	0.71	1.38E-04	0.20	Parts et al.
rs11191560	10	C	Gluteal SAT	SFXN2	0.153	1.72E-05	0.20	rs71496550	NA	4.42E-06	0.41	Min et al.
rs11191560	10	C	Abdominal SAT	SFXN2	0.628	1.44E-04	0.02	rs71496550	NA	9.13E-05	0.94	Min et al.
rs7164727	15	T	Lymphocyte	BBS4	-0.163	3.14E-05	1	rs7164727	1.00	3.14E-05	1	Dixon et al.
rs9925964	16	A	Liver	VKORC1	0.122	4.41E-37	0.84	rs2303223	0.88	3.62E-44	0.05	Zhong et al.
rs9925964	16	A	Subcutaneous	ZNF646	0.017	2.55E-06	1	rs9925964	1.00	2.55E-06	1	Zhong et al.
rs9925964	16	A	Blood	ZNF668	-0.382	1.70E-12	0.48	rs10871454	0.93	1.10E-12	0.26	Emilsson et al.
rs9914578	17	G	Subcutaneous	C17orf13	-0.010	3.01E-06	0.99	rs7225843	0.99	2.86E-06	0.97	Zhong et al.
rs1808579	18	C	SKIN	C18orf8	-0.073	5.74E-10	0.86	rs1788781	0.90	1.67E-10	0.13	Grundberg et al.
rs1808579	18	C	Subcutaneous	C18orf8	-0.014	8.41E-08	1	rs1808579	1.00	8.41E-08	1	Zhong et al.
rs17724992	19	A	Blood	PGPEP1	-0.825	1.60E-40	1	rs17724992	1.00	1.60E-40	1	Emilsson et al.
Previously reported loci												
rs10182181	2	G	Subcutaneous	ADCY3	0.022	7.57E-06	0.69	rs11684619	0.72	8.70E-09	0.05	Zhong et al.
rs2176040	2	A	Omental	IRS1	-0.036	3.74E-09	0.97	rs908252	0.87	3.98E-10	0.47	Zhong et al.
rs13107325	4	T	Liver	SLC39A8	-0.101	1.29E-17	1	rs13107325	1.00	1.29E-17	1	Zhong et al.
rs205262	6	G	Blood	SNRPC	-0.462	9.60E-15	0.58	rs6457792	0.96	9.40E-15	0.55	Emilsson et al.
rs205262	6	G	PBMC	SNRPC	-0.127	3.40E-09	0.03	rs2744943	0.73	3.15E-11	0.12	PBMC meta-analysis
rs205262	6	G	Omental	SNRPC	-0.012	6.64E-06	0.81	rs2814984	0.75	8.03E-07	0.30	
rs3817334	11	T	SKIN	C1QTNF4	-0.051	1.34E-09	0.82	rs7124681	1.00	9.42E-10	0.34	Grundberg et al.
rs3817334	11	T	Subcutaneous	MTCH2	0.044	7.64E-13	0.76	rs12794570	0.76	2.54E-15	0.10	Zhong et al.
rs3817334	11	T	Brain	MTCH2	28.255	7.51E-08	NA	NA	NA	NA	NA	Myers et al.
rs3817334	11	T	FAT	SPI1	-0.090	9.90E-07	0.90	rs10769262	0.70	1.15E-08	1	Grundberg et al.
rs12016871	13	T	PBMC	GTF3A	-0.258	6.68E-34	0.90	rs7988412	0.81	1.81E-36	0.29	PBMC meta-analysis
rs12016871	13	T	Lymphocyte	GTF3A	-0.375	3.89E-15	0.32	rs7988412	0.86	1.32E-15	0.06	
rs12446632	16	G	Omental	IQCK	0.028	2.27E-10	0.83	rs11865578	0.83	4.14E-13	0.14	Zhong et al.
rs12446632	16	G	Liver	IQCK	0.031	5.39E-06	0.74	rs9921401	0.70	3.82E-07	0.20	Zhong et al.
rs3888190	16	A	Blood	APOBR	0.303	2.10E-08	0.68	rs2411453	0.83	1.10E-08	0.25	Emilsson et al.
rs3888190	16	A	PBMC	ATXN2L	0.084	1.04E-04	0.99	rs8049439	0.99	8.59E-05	0.88	PBMC meta-analysis
rs3888190	16	A	SKIN	SBK1	-0.063	1.63E-06	0.41	rs4788084	0.82	2.87E-07	0.10	
rs3888190	16	A	Adipose	SH2B1	-0.407	4.10E-13	0.67	rs12928404	0.92	2.40E-13	0.30	Emilsson et al.
rs3888190	16	A	Omental	SH2B1	-0.014	5.29E-07	0.87	rs12928404	0.93	4.65E-07	0.83	Zhong et al.
rs3888190	16	A	Subcutaneous	SULT1A2	0.067	3.36E-21	0.52	rs1074631	0.80	3.93E-23	0.14	Zhong et al.
rs3888190	16	A	PBMC	TUFM	0.694	9.81E-198	0.94	rs8049439	0.99	9.81E-198	0.12	PBMC meta-analysis
rs1808579	18	C	Subcutaneous	NPC1	-0.027	2.52E-10	0.83	rs1805081	0.78	7.86E-14	0.06	
rs3888190	16	A	SKIN	TUFM	0.074	7.90E-10	0.46	rs2411453	0.76	1.91E-10	0.09	Grundberg et al.
rs3810291	19	A	Adipose	ZC3H4	-0.386	3.70E-09	1	rs3810291	1.00	3.70E-09	1	Emilsson et al.

Extended Data Table 4 | Putative coding variants in LD ($r^2 \geq 0.7$) with GWS BMI loci

BMI SNP	Chr.	Source	Putative Coding Variant	r^2	Gene	Protein Alteration	PhastCon Score	GERP Score	Grantham Score	PolyPhen	SIFT Prediction	SIFT Score
Novel genome-wide significant loci												
rs492400	2	1000G	rs3770213	0.89	<i>ZNF142</i>	L956H	0	-1.6	99	possibly damaging	Damaging	0
rs492400	2	1000G	rs3770214	0.89	<i>ZNF142</i>	S751G	0.2	1.4	56	benign	Tolerated	0.08
rs492400	2	1000G	rs2230115	0.963	<i>ZNF142</i>	A541S	0.5	5.1	99	benign	Tolerated	0.044
rs492400	2	1000G	rs1344642	0.963	<i>STK36</i>	R583Q	0	2.4	43	possibly damaging	Damaging	0
rs492400	2	1000G	rs1863704	0.89	<i>STK36</i>	G1003D	0	2	94	possibly damaging	Tolerated	0.41
rs492400	2	1000G	rs1863704	0.89	<i>STK36</i>	G982D	0	2	94	possibly damaging	-	-
rs492400	2	1000G	rs3731877	0.792	<i>TTLL4</i>	E34Q	1	5.5	29	probably damaging	Unknown	Not scored
rs17001654	4	1000G	rs61750814	1	<i>NUP54</i>	N250S	1	5.5	46	benign	Damaging	0.05
rs4740619	9	1000G	rs4741510	0.901	<i>CCDC171</i>	S121T	1	2	58	benign	Damaging	0.05
rs4740619	9	1000G	rs1539172	0.74	<i>CCDC171</i>	K1069R	1	4.1	26	benign	Tolerated	1
rs2176598	11	1000G	rs11555762	0.774	<i>HSD17B12</i>	S280L	0	0.4	145	benign	Tolerated	0.74
rs3849570	3	1000G	rs2229519	0.771	<i>GBE1</i>	R190G	1	4.8	125	benign	Damaging	0.04
rs3736485	15	1000G	rs12102203	0.966	<i>DMXL2</i>	S1288P	0.7	1.7	74	benign	Tolerated	0.32
rs7164727	15	1000G	rs2277598	0.839	<i>BBS4</i>	I182T	0	-4.4	89	benign	Tolerated	0.47
rs9925964	16	1000G	rs749670	0.869	<i>ZNF646</i>	E327G	1	4.2	98	benign	Tolerated	0.44
Previously identified genome-wide significant loci												
rs10182181	2	HapMap	rs11676272	0.967	<i>ADCY3</i>	S107P	0	2.9	74	benign	Tolerated	0.28
rs13107325	4	1000G	rs13107325	1	<i>SLC39A8</i>	A324T	1	4.4	5.8	benign	Tolerated	0.09
rs13107325	4	1000G	rs13107325	1	<i>SLC39A8</i>	A391T	1	4.4	5.8	benign	Tolerated	0.09
rs2112347	5	1000G	rs2307111	0.862	<i>POC5</i>	H11R	0.9	5.8	29	benign	Unknown	Not scored
rs2112347	5	1000G	rs2307111	0.862	<i>POC5</i>	H36R	0.9	5.8	29	benign	Unknown	Not scored
rs4256980	11	HapMap	rs7935453	0.729	<i>TRIM66</i>	L630V	-	-	-	-	Tolerated	1
rs4256980	11	1000G	rs11042022	0.876	<i>TRIM66</i>	H466R	-	-	-	-	Tolerated	0.38
rs4256980	11	1000G	rs11042023	0.959	<i>TRIM66</i>	H324R	1	5.1	29	probably damaging	Damaging	0.03
rs11030104	11	1000G	rs6265	0.817	<i>BDNF</i>	V148M	1	5.2	21	probably damaging	Damaging	0
rs11030104	11	1000G	rs6265	0.817	<i>BDNF</i>	V66M	1	5.2	21	probably damaging	Damaging	0
rs11030104	11	1000G	rs6265	0.817	<i>BDNF</i>	V74M	1	5.2	21	probably damaging	Damaging	0
rs11030104	11	1000G	rs6265	0.817	<i>BDNF</i>	V81M	1	5.2	21	probably damaging	Damaging	0
rs11030104	11	1000G	rs6265	0.817	<i>BDNF</i>	V95M	1	5.2	21	probably damaging	Damaging	0
rs3817334	11	1000G	rs1064608	0.809	<i>MTCH2</i>	P290A	1	5.1	27	probably damaging	Tolerated	0.12
rs3888190	16	1000G	rs180743	0.789	<i>APOBR</i>	P428A	0.1	0.5	27	benign	Unknown	Not scored
rs3888190	16	1000G	rs7498665	1	<i>SH2B1</i>	T484A	1	3.1	58	benign	Tolerated	0.25
rs16951275	15	1000G	rs7170185	1	<i>LBXCOR1</i>	W200R	-	-	-	-	-	-
rs1808579	18	1000G	rs1805082	0.935	<i>NPC1</i>	I858V	1	6.1	29	benign	Tolerated	0.24
rs1808579	18	1000G	rs1805081	0.905	<i>NPC1</i>	H215R	0	-1.1	29	benign	Tolerated	0.59
rs2287019	19	1000G	rs1800437	0.714	<i>GIPR</i>	E354Q	1	3.1	29	probably damaging	Tolerated	0.09

r^2 is the LD between the BMI index SNP and the putative coding variant.

Shearing-induced asymmetry in entorhinal grid cells

Tor Stensola^{1*}, Hanne Stensola^{1*}, May-Britt Moser¹ & Edvard I. Moser¹

Grid cells are neurons with periodic spatial receptive fields (grids) that tile two-dimensional space in a hexagonal pattern. To provide useful information about location, grids must be stably anchored to an external reference frame. The mechanisms underlying this anchoring process have remained elusive. Here we show in differently sized familiar square enclosures that the axes of the grids are offset from the walls by an angle that minimizes symmetry with the borders of the environment. This rotational offset is invariably accompanied by an elliptic distortion of the grid pattern. Reversing the ellipticity analytically by a shearing transformation removes the angular offset. This, together with the near-absence of rotation in novel environments, suggests that the rotation emerges through non-coaxial strain as a function of experience. The systematic relationship between rotation and distortion of the grid pattern points to shear forces arising from anchoring to specific geometric reference points as key elements of the mechanism for alignment of grid patterns to the external world.

Grid cells of the mammalian medial entorhinal cortex (MEC) are part of a neural metric for self-localization that is independent of contextual details^{1–5}. Grid patterns are thought to arise not by direct extraction of features from sensory inputs, but by local network dynamics within the entorhinal cortex itself^{6–11}. The coherent response of ensembles of grid cells following experimental interventions^{4,11,12}, and the modular organization of the grid cells^{12,13}, point to a neural architecture in which grid patterns emerge by competitive interactions between interconnected groups of neurons and move across the network in accordance with the animal's trajectory in the outside world^{5–11}.

For grid cells to be effective as an internal representation of the animal's location, the grid pattern must be anchored consistently to an external reference frame provided by stationary sensory cues^{2,12,13}. Grid patterns may be anchored at multiple locations and to multiple reference frames, as suggested by the observation that the grid pattern fragments into a mosaic of local sub-grids in environments with multiple compartments¹⁴. However, how the grid pattern is stabilized by specific features of the environment is poorly understood. To address the mechanisms of the anchoring process, we determined the orientation and shape of the grid for a sample of 807 grid cells collected in familiar large square open-field environments¹². Grid orientation could be determined accurately because 80% of these cells (469/587 cells) had 6 or more grid fields. Square recording environments were chosen because they contain axes of symmetry that are distinct from the symmetry axes of the grid pattern and so might be suitable for inferring solutions to the alignment problem.

Grid orientation reflects boundaries

The orientation of the grid axes did not distribute randomly across modules and animals. In each module of grid cells from each animal, the orientation of the grid axes clustered around the cardinal axes of the environment, defined by the north–south and east–west walls of the test box, confirming previous work showing that axes of the grid pattern tend to align with the walls in a square environment¹⁵.

The majority of data were recorded in a 1.5 m wide square enclosure (7 data sets, 6 rats, 587 grid cells, 18 grid modules). In this environment,

grid orientations clustered around the east–west axis and 60° multiples of this axis (Fig. 1a–c, mean \pm s.e.m. grid orientation averaged across the three main grid axes: $-1.91 \pm 0.24^\circ$). Clustering around these axes was observed for all grid modules in all animals (Fig. 1b, c and Extended Data Figs 2 and 3). A Rayleigh test confirmed that mean grid orientation, averaged across grid axes, was highly non-uniform ($Z = 405$, $P = 2.2 \times 10^{-226}$, distribution multiplied by 6 to achieve 360°).

However, although the orientation of the grid pattern distributed symmetrically around the east–west axis, closer inspection suggested that most grid cells were offset by a few degrees in either direction from the cardinal axis (Fig. 1c, d). To quantify the offset, we identified for each grid axis (Ax1–Ax3) in each cell the wall that had the smallest angular deviation from that axis (denoted W_{N1-N3}). We then selected the minimal value A_{\min} among the three angles (Extended Data Fig. 1a). The distribution of A_{\min} across the cell population had a sharp unimodal peak centred at 7.40° (mean absolute value \pm s.e.m.: $7.15 \pm 0.15^\circ$; Fig. 1d), not far off from the 7.5° offset that would have minimized symmetry or overlap with the cardinal and diagonal axes of the environment (Fig. 1e, f). For the vast majority of grid cells in the 1.5 m environment (575/587 cells), A_{\min} was defined by the east–west axis (average positive offset: $7.48 \pm 0.17^\circ$, $n = 362$; negative offset: $6.11 \pm 0.22^\circ$, $n = 213$). Very few data points distributed near 0° and 15° , the two orientations that overlapped with the symmetry axes of the box (Fig. 1e, f; Extended Data Fig. 1b). The distribution of A_{\min} was highly non-uniform (Rayleigh test, $Z = 62.5$, $P = 1.4 \times 10^{-28}$).

A_{\min} was similar between cells that belonged to the same grid module, but the direction often differed across modules (Extended Data Fig. 4). Positive and negative offsets distributed across animals with nearly equal probability (7/18 modules had positive offsets). In animals with data from ≥ 2 modules, the offset directions appeared to be drawn independently (35% of module pairs within animals had different offset directions; $P = 0.26$, binomial test), suggesting that grid orientation is determined within modules and not globally as a function of behaviour.

The rotational offset differed between the three grid axes. For a perfectly hexagonal grid, as a result of sixfold symmetry, a 7.5° offset in axis 1 should result in distributions around 22.5° from W_{N2} and 37.5°

¹Kavli Institute for Systems Neuroscience and Centre for Neural Computation, Norwegian University of Science and Technology, Olav Kyrres gate 9, 7491 Trondheim, Norway.

*These authors contributed equally to this work.

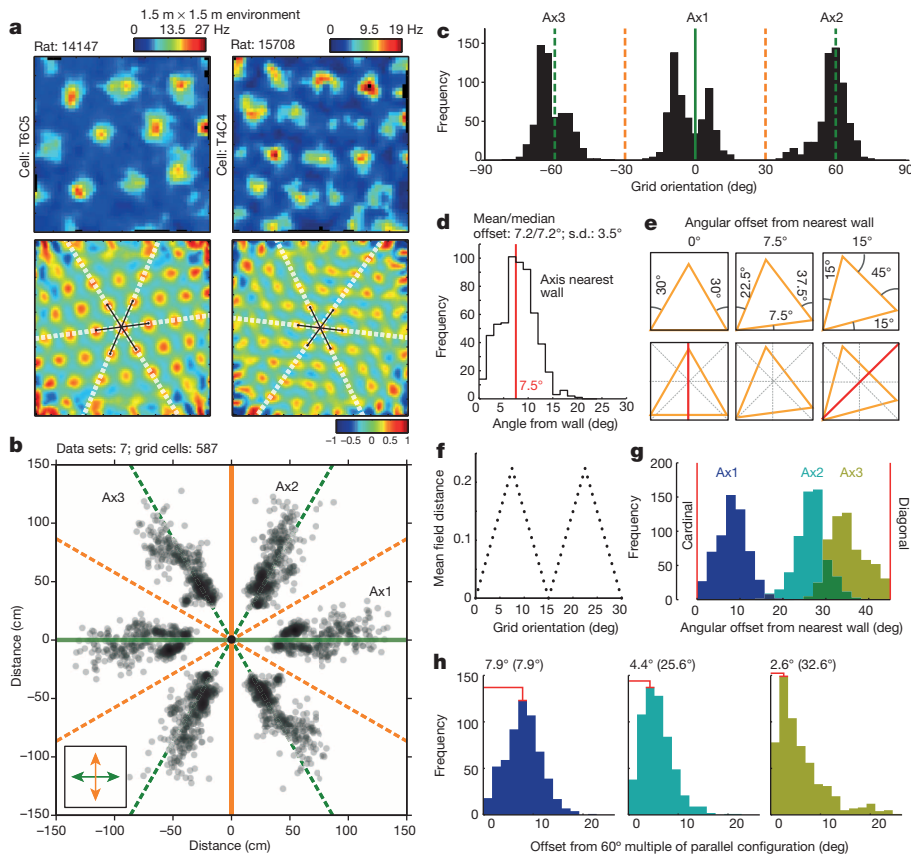


Figure 1 | Grid alignment in a 1.5 m square environment. **a**, Colour-coded firing-rate maps (top) and spatial autocorrelograms (bottom) for representative grid cells from two animals in the 1.5 m square. Innermost hexagon of vertices is indicated. **b**, Polar scatterplot showing distribution of grid orientation and grid spacing for 587 grid cells in the 1.5 m box. For each cell, the location of the 6 innermost fields in the spatial autocorrelogram is shown (6 dots per cell; distance from the centre is proportional to grid scale; box cardinal axes in solid orange (north–south) and solid green (east–west); dashed coloured lines show 60° offsets from cardinal axes). **c**, Frequency distributions showing clustering of data in **b** around 60° multiples of east–west orientation. Axis 1 is east–west. **d**, Distribution of orientation for grid axis with smallest offset from box walls (axis 1, A_{\min}). Red line indicates orientation that would provide minimal symmetry (7.5°, Fig. 1e, f). **e**, Top, schematic showing expected offsets for individual grid axes of a perfectly hexagonal grid at different degrees of rotation. Bottom, overlap between symmetry axes of box and grid at these rotations. Red lines show common symmetry axes. **f**, Minimal distance between field locations in a hexagonal grid and in the same pattern reflected around the cardinal or diagonal axes of the box, estimated for grid orientations in 0.5° steps. **g**, Distribution of angular offsets from the nearest wall axis W_N for each grid axis. **h**, Distributions of angular offset from W_N or 60° multiples of this orientation. Peak angular offsets from nearest 60° multiple are indicated. Offset from respective W_N in brackets.

from W_{N3} for axis 2 and 3, respectively (Fig. 1g, h; Extended Data Fig. 1b). However, in the data, the greater the angle between a grid axis and the corresponding W_N , the smaller was the deviation from the nearest 60° multiple of the parallel solution ($A_{\min} = 0^\circ$) (Fig. 1h, Extended Data Fig. 1d–g). As a consequence of the distinct rotations, the hexagonal symmetry of the grid pattern was perturbed. In most cells the inner 6 fields of the spatial autocorrelogram formed an ellipse rather than a circle¹² (Fig. 2a, b; Extended Data Fig. 5). A similar deformation of the grid pattern has been observed under other experimental circumstances^{12,13,16}.

Grid rotation reflects shearing

In continuum mechanics, shearing is a deformation where points on a plane are shifted along the shear axis with a displacement proportional to the distance from the shear axis¹⁷ (Fig. 2c). Distances between points that are perpendicular to the shear axis remain constant. The shear transform is described for two dimensions by:

$$f(x, y) = \begin{bmatrix} 1 & \gamma_1 \\ \gamma_2 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

or, with algebraic notation:

$$f(x, y) = \begin{bmatrix} x + \gamma_1 y \\ y + \gamma_2 x \end{bmatrix}$$

where γ_1 is the shear parameter along the y axis, γ_2 the shear parameter along the x axis, and x and y are row vectors of initial coordinates of points in the plane. For simple shearing, only one of the shear parameters has a non-zero value. When the points that are deformed by simple shearing lie on a circle, the circle becomes elliptical (with the resulting ellipse referred to as the strain ellipse). This deformation is accompanied by non-coaxial rotation of axes in this circle, proportional to the angular distance from the shear axis. Forces that act on a plane to produce simple shearing can be deduced from the ellipticity and orientation (the semi-major and semi-minor axes) of the strain ellipse.

The close link between deformation and rotation of the grid pattern led us to hypothesize that elliptification of the grid reflects shearing along either of the box axes and that, through non-coaxial rotation, such deformations underlie the consistent angular offset of the grid axes (Fig. 2c). We tested these hypotheses by applying simple shear transformations to all grid patterns, with shear axes parallel to the box axes, in the reverse direction of the transform that might have caused the rotation in the first place (Fig. 2d). For each cell, we determined the shear parameter that reversed grid ellipticity to a minimum along either cardinal axis.

The effects were direction-dependent. Shearing along the north–south axis, orthogonal to the alignment axis (east–west), reduced ellipticity from 1.17 ± 0.004 to 1.06 ± 0.009 (means \pm s.e.m.). The angular offset of the grid pattern was completely abolished by this transform (Fig. 2d, e). After shearing, orientations distributed unimodally around the east–west axis, with almost no offset (peak for A_{\min} at -0.9° , s.d.: 6.6° ; A_{\min} before versus after shearing: $Z = 10.7$, $P = 7.4 \times 10^{-27}$, Wilcoxon rank sum test). There was a strong correlation between the angular offset of the grid and the optimal shear parameter γ , that is, the force used to minimize ellipticity in the grid (Fig. 2f; $r = -0.60$, $P = 1.04 \times 10^{-58}$; γ versus absolute angular offset: $r = 0.16$, $P = 5.34 \times 10^{-4}$). Mere compression towards the diagonal could not reproduce the relationship between deformation and offset (Extended Data Fig. 6). Shearing in the east–west direction, parallel to the alignment axis, did not reduce the offset of the grid, despite similar attenuation of the ellipticity (Fig. 2e; $Z = 0.44$, $P = 0.66$; ellipticity \pm s.e.m.: 1.06 ± 0.009). Grid axes with a smaller deviation from the shear axis showed less rotation (Fig. 1g, h; Extended Data Fig. 1d–g). These observations would all be expected if the grid axes exhibit shearing-induced non-coaxial rotation, proportional to their angular distance from the shear axis.

Finally, we tested, in a different set of rats, whether shearing developed with experience, as hypothesized. We found that when rats explored a square box for the first time, the grid was only minimally

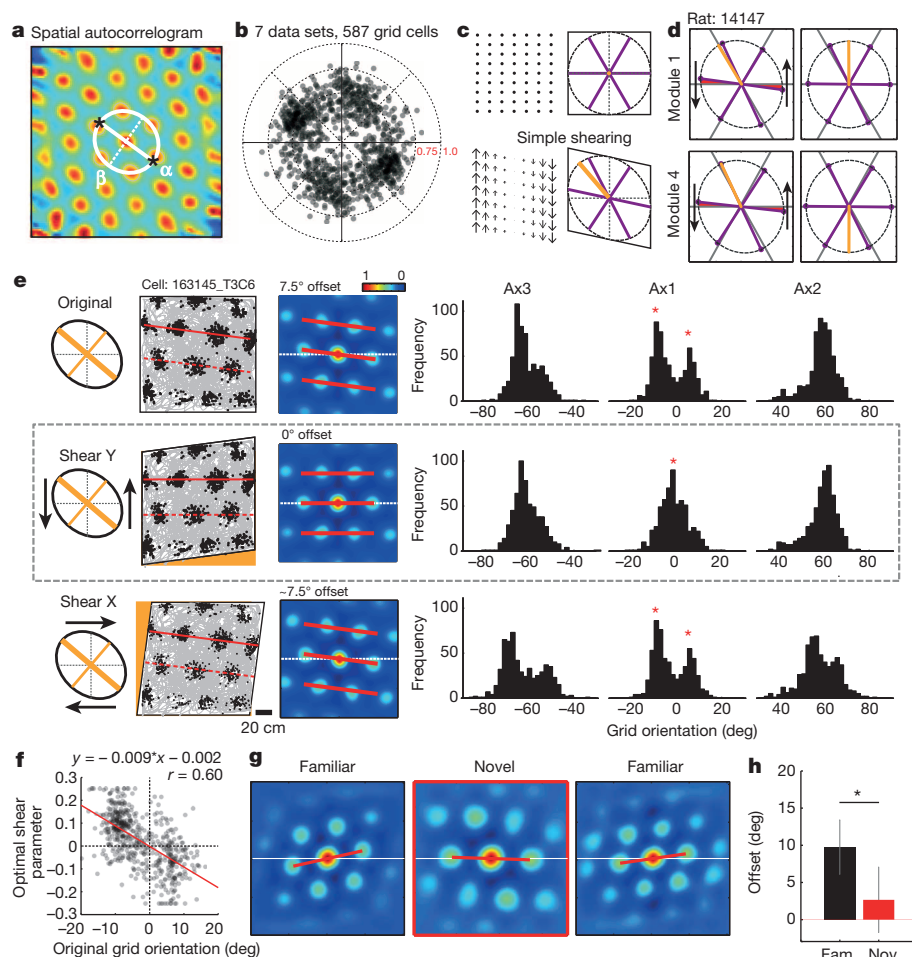


Figure 2 | Relationship between orientation and deformation of grid patterns. **a**, Autocorrelogram showing ellipse fit to inner hexagon of firing fields of an example cell, with semi-major and semi-minor axes indicated as α and β . **b**, Polar scatterplot showing eccentricity and orientation of ellipse fits for each cell in a sample of 587 grid cells in the 1.5 m environment (two dots per cell, as indicated with asterisks in **a**). **c**, Schematic showing deformation and non-coaxial rotation before and after simple shearing. Amplitude and direction of shear forces are indicated by arrows (bottom left). Bottom right, orientation of strain ellipse (orange line) and effect on grid orientation (purple axes). **d**, Schematic showing reversal of grid orientation in two example ellipses (modules M1 and M4) following shearing in the indicated direction. Minimizing ellipticity by reverse north-south shearing is accompanied by re-alignment of grid axes. **e**, Left, spike maps (trajectories with spikes) and autocorrelograms showing grid orientation for an example cell (module M1) before and after reverse shearing along each cardinal axis. A_{\min} is indicated by red lines. Right, distribution of orientation for each grid axis following minimization of ellipticity by reverse shearing. Note disappearance of 7.5° orientation peaks (asterisks) only following north-south shearing. **f**, Scatterplot showing strong correlation between angular offset (A_{\min} before shearing) and optimal shear parameter, that is, amount of shearing required to minimize ellipticity. **g**, Autocorrelation maps for a representative grid cell from trials in familiar and novel boxes (from left to right: first trial familiar, novel, second trial familiar). Red lines indicate A_{\min} . **h**, Mean orientation offset (\pm s.e.m.) for cells recorded in both environments. $*P < 0.05$.

offset from the wall (Fig. 2g, h; Extended Data Fig. 7; 20 cells, median A_{\min} of 2.7°). A_{\min} was significantly larger when the same animals explored a familiar square (median of 9.8° , 13 cells, all recorded also in the novel box, Wilcoxon rank sum test: $Z = 2.12$, $P < 0.05$), suggesting that when an animal encounters an environment for the first time, one of the grid axes is aligned to one of the walls and, over time, shearing causes the axes to rotate away in an orientation-dependent manner.

Shear forces may operate interactively

In large environments, the ability of the grid map to maintain accurate spatial representations may be jeopardised by the increased distance between salient geometric features. To address this possibility, we analysed grid orientation and deformation in 220 cells (13 modules, 5 rats) from a 2.2 m wide square box, with a surface area more than twice that of the 1.5 m environment. In this environment, the cells clustered around both cardinal axes (Fig. 3a–c; Rayleigh test, $Z = 18.9$, $P = 4.4 \times 10^{-9}$). As in the 1.5 m box, A_{\min} was not far off from 7.5° (Fig. 3d; mean of 7.37° ; Extended Data Figs 1h and 4). Only a few cells had offsets near 0° or 15° . A small cell sample was also tested in a large circular environment (2.0 m diameter; Extended Data Fig. 8). Here grid orientation was more distributed, as expected in the absence of unique reference orientations.

We asked if grids in the 2.2 m box underwent shearing-induced deformation in the same way as in the smaller environment, and if grid orientation offsets still could be eliminated by reverse shearing. Grid cells were as elliptical as in the smaller environment (mean \pm s.e.m. ellipticity: 1.17 ± 0.007), but the distribution of ellipse orientations was less clustered (Fig. 3e, Extended Data Fig. 5d–g; Rayleigh test: $Z = 2.97$, $P = 0.05$), suggesting that, in the large box, the deformation might not be ascribed to shearing along a single cardinal axis. To test this possibility,

we again sheared each grid pattern along either of the box axes. Separate analyses were performed for cells that aligned to the north-south and east-west axes. Simple single-axis shearing reduced grid ellipticity to a similar extent as in the smaller environment (from 1.17 ± 0.007 to 1.06 ± 0.007) but the angular offset of the grid (A_{\min}) was not fully eliminated, even when the shear axis was orthogonal to the alignment axis (mean offset \pm s.e.m. from $7.4 \pm 0.02^\circ$ to $4.9 \pm 0.02^\circ$; $Z = 6.4$, $P = 1.6 \times 10^{-10}$, Wilcoxon rank sum test; Fig. 3f; Extended Data Fig. 9a, b). The offset was not reduced when the shear axis was parallel to the alignment axis (mean \pm s.e.m. after shearing: $8.6 \pm 0.02^\circ$).

The fact that shearing only partly removed the angular offset raises the possibility that anchoring was maintained also by forces from the orthogonal wall. Indeed, the rotational offset of the 3 grid axes in the 2.2 m box did not match predictions from single axis shearing but fit well with added shearing from the second wall (Fig. 3f, g, Extended Data Figs 5h and 9c–e). When shearing from 2 axes was applied simultaneously across a range of shear parameter values, unique solutions could be found that minimized ellipticity to a much better extent than in the one-axis scenario, with complete reversal of the angular offset in cases where the grid was anchored to diagonally opposite corners (Fig. 3h; Extended Data Figs 7 and 10).

Shear forces may operate locally

Consistent with the idea that grids anchor locally, some grid modules had orientations in the 2.2 m box that changed gradually (Fig. 4a, c) or discretely (Fig. 4b, d) across the arena. In cells with heterogeneous anchoring patterns, the angular offsets of fragments of the grid, but not the global orientation, remained close to 7.5° (Fig. 4c, d). Simple combinations of shear-like forces from particular walls could reproduce the graded and fragmented patterns in simulated grids, with orientation

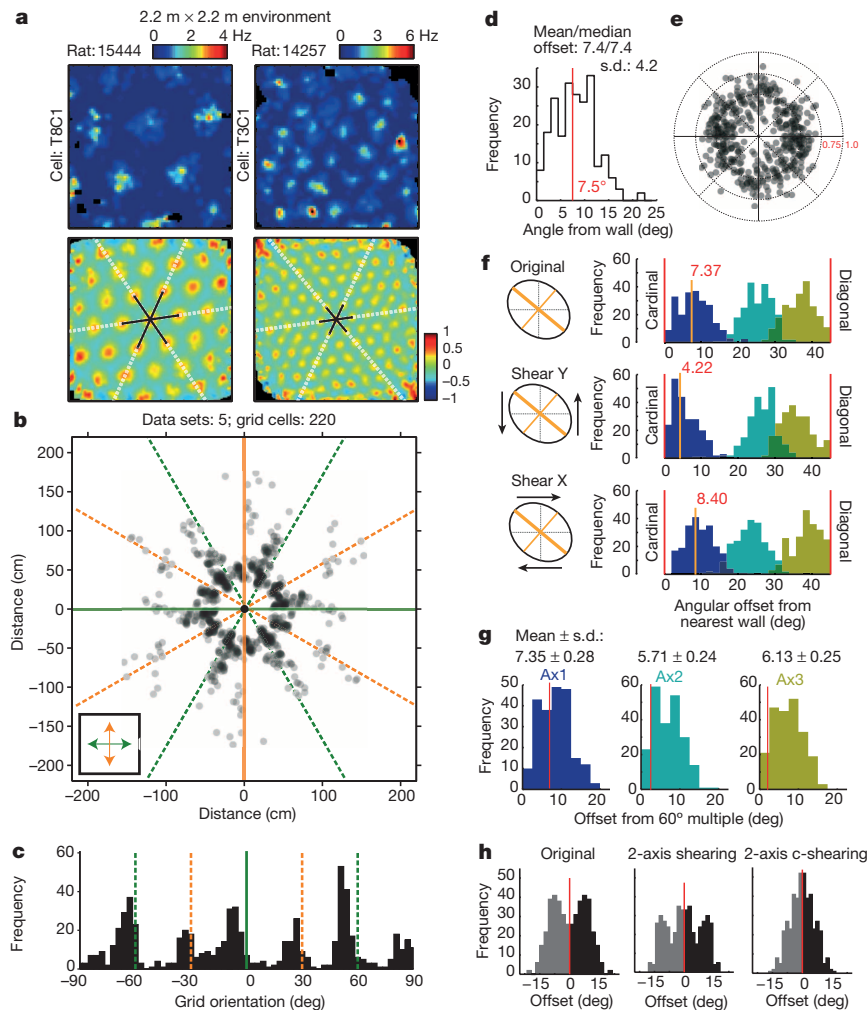


Figure 3 | Grid alignment in a 2.2 m square environment. **a**, Rate maps and spatial autocorrelograms for representative grid cells from the 2.2 m square. Symbols as in Fig. 1a. **b**, Polar scatterplot showing distribution of grid orientation and grid spacing for 220 grid cells in the 2.2 m box. **c**, Frequency distributions showing clustering around 60° multiples of the two cardinal axes. **d**, Distribution of A_{\min} , as in Fig. 1d. **e**, Polar scatterplot showing distribution of ellipse fits as in Fig. 2b. **f**, Distributions of grid orientations following minimization of ellipticity by shearing along either cardinal axis. Axis distributions as in Fig. 1g. Cells aligned to the east-west walls were first rotated 90°. **g**, Frequency distributions showing, for each grid axis, the angular offset from the nearest 60° multiple of the cardinal axes (as in Fig. 1h). Distribution means and s.d. are indicated. Red lines show the predicted offsets following simple shearing orthogonal to the alignment axis until A_{\min} is 7.5°. Note systematic deviations in axis 2 and axis 3, suggesting shearing from a second wall had taken place. **h**, Histogram of angular offsets after two forms of two-axis shearing (all data from 2.2 m box). Black bars show data; grey bars show reflections around 0°. For each grid, we explored two-axis shear parameter space and determined the unique solution for minimal ellipticity. Left, original distribution of grid offset. Middle, distribution after two-axis standard shearing. Right, distribution after shearing with anchoring to each of two diagonally opposite corners (Extended Data Figs 7 and 10).

offsets and ellipse orientations that matched observed values (Fig. 4c, d; Extended Data Fig. 9d, e). This suggests that the same forces operate on locally anchored grid segments as on grids with uniform global geometry.

In order to evaluate the extent of geometric heterogeneity in the grid pattern following local alignment, we estimated deformation and orientation of grid cells in each quadrant of the 1.5 m and 2.2 m environments (Fig. 4e). The difference in grid properties across quadrants was determined by calculating the product of all pairwise correlations between module-specific autocorrelograms made for each quadrant (Fig. 4e–h and Extended Data Fig. 9f). In the 1.5 m enclosure, this product was large (Fig. 4f; median ± s.d.: 0.81 ± 0.11 , 9 modules, 3 animals), indicating consistent global alignment. In the 2.2 m enclosure, the product was significantly lower (Fig. 4g, h; median ± s.d.: 0.60 ± 0.21 , 8 modules, 2 animals; Wilcoxon rank sum test, rank sum: 105, $P = 0.02$), suggesting that grid patterns were less uniformly anchored.

To determine if alignment to borders is distance-dependent, we finally asked whether deformation of the grid pattern is stronger near walls and corners than in central parts of the box. We divided the data from the 2.2 m box into 9 (3×3) sub-divisions (Fig. 4i). Grid orientation and grid deformation were then estimated for each grid module with 6 or more detectable fields in each sub-division. The central bin of the 2.2 m box showed significantly higher grid scores across modules and animals than the corner bins (Fig. 4j; Kruskal–Wallis ANOVA: $H = 6.6$, d.f. = 2, $P < 0.05$, Tukey post-hoc test). The decrease in rotational symmetry at the corners was accompanied by a significant increase in ellipticity in these bins compared to the central bin (Fig. 4k; $H = 7.99$, d.f. = 2, $P < 0.05$), suggesting that distortions increase with proximity to the corners. One particular corner showed a remarkably

low variation in the distribution of ellipse orientation (Fig. 4l). For every single animal, this corner segment was where the animal was released into the box, raising the possibility that anchoring is determined by the animal's initial experience.

Discussion

The findings provide evidence for a mechanism by which geometric features of the environment cause local rotation and deformation of the hexagonal firing patterns of grid cells. Previous work has suggested that grid cells are oriented with one axis more or less parallel to one of the walls in a square environment¹⁵. Our findings confirm this observation but show that within the scatter of grid orientations there is a limited number of orientation solutions, each corresponding to an average offset of $\sim 7.5^\circ$ in either direction from either wall of the environment. We show that the rotation of the grid axes is accompanied by elliptic deformation of the grid pattern, raising the possibility that rotation and deformation are part of a single process.

In the small environment, the orientation offset could be completely and selectively abolished by non-coaxial rotation through a simple shear transform that minimized distortion (ellipticity) of the grid pattern specifically along one of the wall axes, in the reverse direction of the rotation that developed as animals got familiar with the environment. In the large environment, the grid pattern aligned interactively and in a distance-dependent manner to several walls. However, minimization of ellipticity by a two-axis shearing transform abolished the offset. The findings suggest that the rotational offset of the grid pattern is a consequence of shearing forces determined by the symmetry axes of the environment, and that, for a specific axis of the grid,

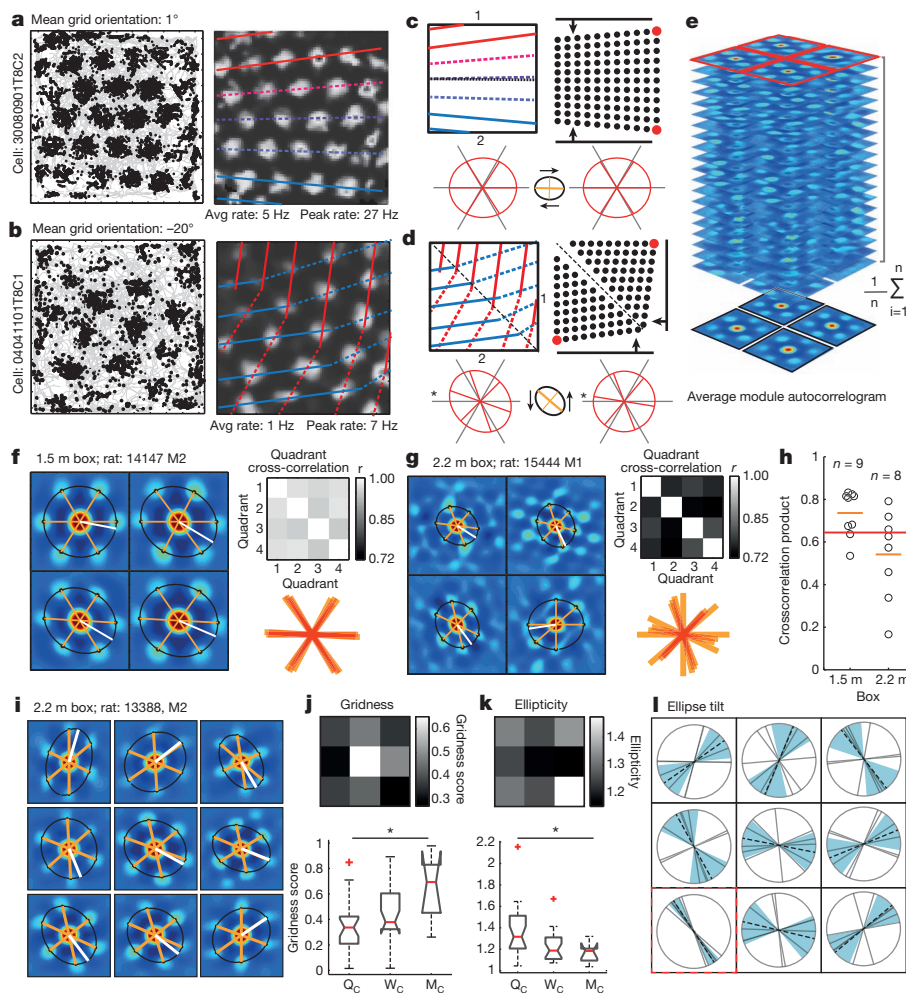


Figure 4 | Grids can have multiple alignment solutions. **a, b,** Spike and rate maps for two grid cells in the 2.2 m box. Left, spikes (black) superimposed on trajectory (grey). Right, corresponding brightness-coded rate maps (white is maximum). Coloured lines illustrate change of orientation. **c, d,** Top left panels, summary of change in A_{\min} across the test arena. Symmetry axes are shown as dashed black lines. Top right panels, square lattices after shear-like interactions from two walls (black lines with arrows). Anchoring points are shown by red dots. Bottom panels, grid axes and ellipses calculated from patterns in top right panels before (left) and after (right) shearing to minimize ellipticity. Asterisk, offset is still present. **e,** Stack of spatial autocorrelograms made from box quadrants (Extended Data Fig. 9f; rat: 15444, module M2). **f, g,** Colour-coded mean autocorrelograms for each quadrant, using maps from all grid cells in a module (as in **e**) in one data set from the 1.5 m box (**f**) or the 2.2 m box (**g**). Ellipse fits (black), semimajor axis (white), and grid axes (orange) are indicated. Top right, brightness-coded matrices showing cross-correlation of autocorrelograms for all combinations of quadrants. Bottom right, superimposed grid axes from all 4 autocorrelograms. **h,** Product of all pairwise crosscorrelations for individual experiments (circles). **i,** Mean autocorrelograms for each of 3×3 subdivisions of the 2.2 m environment, with maps constructed as in **e**. Ellipses and A_{\min} were fitted as in **f, g**. **j,** Gridness scores for individual subdivisions across data sets. Top, brightness-coded matrix showing mean gridness for each bin. Bottom, boxplot of gridness scores for the 4 corner bins (Q_c), the 4 remaining wall bins (W_c), and the single centre bin (M_c). * $P < 0.05$. **k,** Same as in **j**, but for grid ellipticity. * $P < 0.05$. **l,** Distribution of ellipse orientation (grey lines) across subdivisions. Dashed black lines indicate circular mean orientations; blue areas show circular s.d.

the angular rotation depends directly on the angular deviation from the shear axis.

How shearing is implemented in the entorhinal microcircuit has not yet been established. Shearing may begin when grid fields are anchored by associative plasticity to features of the boundaries, such as corners, during the animal's first exposure to the environment and this initial anchoring is followed on subsequent sessions by shrinkage of the grid pattern^{13,18}. With gradual compression and maintained anchoring, the end result is a deformation of the grid, accompanied by rotation of those grid axes that have a strong offset from the shear axis. A related possibility is that grid cells are subject to continuous repulsion from the boundaries of the environment¹⁵. The rotational offset of the grid speaks against a uniform repulsion, but repulsive forces from the walls may lead to shearing of the grid pattern if the grid is initially anchored to one or a few reference positions only. Possible mediators of such repulsive effects include entorhinal 'border cells', whose firing fields line up in parallel with the walls and the corners of the environment^{19,20}.

The stabilization of grid orientation at a similar solution of 7.5° in nearly all animals raises the possibility that this orientation confers a functional advantage. This particular rotation minimizes symmetry between the grid pattern and the geometry of the environment. In tandem with the deformation of the grid, the rotation may increase differences between activity patterns at geometrically equivalent locations and thereby reduce the frequency of errors in self-localization, such as the confounding of diagonally opposite corners in rectangular boxes²¹. The observations are consistent with a broad literature pointing to environmental symmetry axes as major determinants of firing locations

in place cells and grid cells^{12,19,20,22–24}, but take these insights further by introducing a mechanism by which forces from the borders align grids to the environment in a way that may optimize the uniqueness of geometrically comparable places.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 17 July; accepted 11 December 2014.

- Fyhn, M., Molden, S., Witter, M. P., Moser, E. I. & Moser, M.-B. Spatial representation in the entorhinal cortex. *Science* **305**, 1258–1264 (2004).
- Hafting, T., Fyhn, M., Molden, S., Moser, M.-B. & Moser, E. I. Microstructure of a spatial map in the entorhinal cortex. *Nature* **436**, 801–806 (2005).
- Sargolini, F. *et al.* Conjunctive representation of position, direction, and velocity in entorhinal cortex. *Science* **312**, 758–762 (2006).
- Fyhn, M., Hafting, T., Treves, A., Moser, M.-B. & Moser, E. I. Hippocampal remapping and grid realignment in entorhinal cortex. *Nature* **446**, 190–194 (2007).
- Moser, E. I. *et al.* Grid cells and cortical representation. *Nature Rev. Neurosci.* **15**, 466–481 (2014).
- Fuhs, M. C. & Touretzky, D. S. A spin glass model of path integration in rat medial entorhinal cortex. *J. Neurosci.* **26**, 4266–4276 (2006).
- McNaughton, B. L., Battaglia, F. P., Jensen, O., Moser, E. I. & Moser, M.-B. Path integration and the neural basis of the 'cognitive map'. *Nature Rev. Neurosci.* **7**, 663–678 (2006).
- Burak, Y. & Fiete, I. R. Accurate path integration in continuous attractor network models of grid cells. *PLOS Comput. Biol.* **5**, e1000291 (2009).
- Couey, J. J. *et al.* Recurrent inhibitory circuitry as a mechanism for grid formation. *Nature Neurosci.* **16**, 318–324 (2013).
- Pastoll, H., Solanka, L., van Rossum, M. C. & Nolan, M. F. Feedback inhibition enables θ -nested γ oscillations and grid firing fields. *Neuron* **77**, 141–154 (2013).
- Yoon, K. *et al.* Specific evidence of low-dimensional continuous attractor dynamics in grid cells. *Nature Neurosci.* **16**, 1077–1084 (2013).

12. Stensola, H. *et al.* The entorhinal grid map is discretized. *Nature* **492**, 72–78 (2012).
13. Barry, C., Hayman, R., Burgess, N. & Jeffery, K. J. Experience-dependent rescaling of entorhinal grids. *Nature Neurosci.* **10**, 682–684 (2007).
14. Derdikman, D. *et al.* Fragmentation of grid cell maps in a multicompartment environment. *Nature Neurosci.* **12**, 1325–1332 (2009).
15. Krupic, J., Bauza, M., Burton, S., Lever, C. & O'Keefe, J. How environment geometry affects grid cell symmetry and what we can learn from it. *Phil. Trans. R. Soc. Lond. B* **369**, 20130188 (2014).
16. Krupic, J., Burgess, N. & O'Keefe, J. Neural representations of location composed of spatially periodic bands. *Science* **337**, 853–857 (2012).
17. Mase, G. *Continuum Mechanics* 44–53 (McGraw-Hill Professional, 1970).
18. Barry, C., Ginzberg, L. L., O'Keefe, J. & Burgess, N. Grid cell firing patterns signal environmental novelty by expansion. *Proc. Natl Acad. Sci. USA* **109**, 17687–17692 (2012).
19. Solstad, T., Boccara, C. N., Kropff, E., Moser, M. B. & Moser, E. I. Representation of geometric borders in the entorhinal cortex. *Science* **322**, 1865–1868 (2008).
20. Savelli, F., Yoganarasimha, D. & Knierim, J. J. Influence of boundary removal on the spatial representations of the medial entorhinal cortex. *Hippocampus* **18**, 1270–1282 (2008).
21. Cheng, K. A purely geometric module in the rat's spatial representation. *Cognition* **23**, 149–178 (1986).
22. O'Keefe, J. & Burgess, N. Geometric determinants of the place fields of hippocampal neurons. *Nature* **381**, 425–428 (1996).
23. Barry, C. *et al.* The boundary vector cell model of place cell firing and spatial memory. *Rev. Neurosci.* **17**, 71–97 (2006).
24. Lever, C., Burton, S., Jeewajee, A., O'Keefe, J. & Burgess, N. Boundary vector cells in the subiculum of the hippocampal formation. *J. Neurosci.* **29**, 9771–9777 (2009).

Acknowledgements We thank M. Mørreaunet, T. Solstad, B. Dunn, Y. Roudi, A. Treves and A. Witoelar for discussions. The work was supported by two Advanced Investigator Grants from the European Research Council ('CIRCUIT', grant agreement number 232608; 'GRIDCODE', grant agreement number 338865), the European Commission's FP7 FET Proactive programme on Neuro-Bio-Inspired Systems (grant agreement 600725), the Kavli Foundation, the Louis-Jeantet Prize for Medicine, and the Centre of Excellence funding scheme of the Research Council of Norway (Centre for the Biology of Memory, grant number 145993; Centre for Neural Computation, grant number 223262).

Author Contributions T.S., H.S., M.-B.M. and E.I.M. planned experiments and analyses, H.S. collected data, T.S. performed analyses, and T.S. and E.I.M. wrote the paper, with inputs from the other authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to T.S. (tor.stensola@ntnu.no) or E.I.M. (edvard.moser@ntnu.no).

METHODS

Source of data and definition of grid cells. The data are taken from two previous publications from our laboratory in which alignment of the grid pattern was not addressed^{2,12}. The 807 cells recorded in familiar square boxes were obtained from ref. 12; the 23 cells recorded in novel square boxes or circular boxes were taken from ref. 2. Grid cells were defined as cells with rotational symmetry scores ('grid score') exceeding the 95th percentile of a shuffled distribution, and modules of grid cells were identified by a *k*-means clustering procedure¹². Mean grid score (\pm s.e.m.) was 0.96 ± 0.009 (95th percentile chance level: 0.20). The data were recorded across a wide area of dorsomedial MEC and contained a range of grid scales, with grid spacing spanning from 35 to 172 cm (mean range of grid spacing per animal \pm s.e.m.: 58.1 ± 33.8 cm).

Subjects. Neural activity was recorded in 23 male Long Evans rats that were 3–5 months old (350–450 g) at the time of implantation (12 from ref. 12; 11 from ref. 2). In 5 of the rats from ref. 12, and in all rats from ref. 2, the tetrodes were implanted near-tangentially to the MEC surface. The remaining 7 rats had multisite implants of vertically oriented tetrodes that collectively covered a large part of the MEC. After surgery, the rats were housed individually in large Plexiglas cages ($45 \times 44 \times 30$ cm) in a humidity and temperature-controlled environment. They were kept on a 12 h light/12 h dark schedule. All testing occurred during the dark phase. The experiments were performed in accordance with the Norwegian Animal Welfare Act and the European Convention for the Protection of Vertebrate Animals used for Experimental and Other Scientific Purposes. The study contained no randomization to experimental treatments and no blinding. Sample size (number of animals) was set a priori to 5 or more, considered as the minimum required for statistical power in parametric tests for the type of data used in the present study.

Electrode implantation and surgery. Tetrodes were constructed from four twisted 17 μ m polyimide-coated platinum-iridium (90–10%) wires (California Fine Wire, CA). The electrode tips were plated with platinum to reduce electrode impedances to between 120–300 k Ω at 1 kHz.

Anaesthesia was induced by placing the animal in a closed glass box filled with isoflurane vapour. Following this, the animal received an i.p. injection of Equithesin (pentobarbital and chloral hydrate; 1.0 ml per 250 g body weight) or anaesthesia was induced by isoflurane (air flow was kept at 1.2 l per minute with 0.5–3% isoflurane as determined by physiological monitoring). For Equithesin anaesthesia, supplementary doses were given when breathing patterns changed and reflexes began to return (0.15 ml per 250 g). Local anaesthetic (Xylocain) was applied on the skin before making the incision. Holes were drilled on the dorsal skull, anterior to the transverse sinus, after which the rat was implanted with either two microdrives carrying a single bundle of 4 tetrodes each (one per hemisphere) or a 'hyperdrive' consisting of 14 independently movable tetrodes. Hyperdrive implants were always on the left side, whereas microdrives were implanted on both sides. Microdrive tetrodes were inserted at an angle of 20° relative to the bregma/lambda horizontal reference, with the tips pointing in the anterior direction. The tetrodes were inserted 0.1 mm anterior to the transverse sinus edge. Hyperdrive tetrodes were implanted vertically between 3.75 and 5.0 mm lateral to bregma, with the posterior border of the bundle located 0.1–0.2 mm anterior to the edge of the transverse sinus. Jewellers' screws and dental cement were used to secure the drive to the skull. One or two screws in the skull were connected to the drive grounds. During the first days after the surgery, the animals received oral doses of the analgesic Metacam (Meloxicam, 0.1 mg per 300 g; Boehringer Ingelheim, Germany).

Recording procedures. Over the course of ~1–3 weeks, tetrodes were lowered in steps of 300 μ m or less until large-amplitude theta-modulated activity appeared in the local field potential at a depth of about 2.0 mm or lower. In hyperdrive experiments, two of the tetrodes were used, respectively, to record a reference signal from white matter areas. The drive was connected to a multichannel, impedance matching, unity gain headstage. The output of the headstage was conducted via a counterbalanced lightweight multiwire cable to an Axona acquisition system (Axona Ltd, Herts, UK, for all tangentially implanted animals) or via a lightweight multiwire tether cable and through a slip-ring commutator to a Neuralynx data acquisition system (Neuralynx, Tucson, AZ; Neuralynx Cheetah 64, for all multisite implanted animals). Both cables allowed the animal to move freely within the available space. Unit activity was amplified by a factor of 3,000–5,000 and band-pass filtered from 800 to 6,700 Hz (Axona) or from 600 to 6,000 Hz (Neuralynx). Spike waveforms above a threshold set by the experimenter (~55 μ V) were time-stamped and digitized at 32 kHz for 1 ms. EEG signals were amplified by a factor of 1,000 and recorded continuously between 0 and 475 Hz at a sampling rate of 1,893 Hz or 2,034 Hz. Light-emitting diodes (LEDs) on the headstage were used to track the animal's movements at a sampling rate of 25 Hz (Neuralynx) or 50 Hz (Axona). The rat rested on a towel in a large flowerpot on a pedestal while electrical activity was monitored.

Data from the novel environment were recorded after the animals had been trained for 12 days or more in an accompanying familiar environment. Familiar trials lasted 10 min; novel trials lasted 10–30 min, until the novel space was fully

covered by the animal. Stable grid patterns could be seen from the beginning in the novel box (mean gridness \pm s.e.m. for the entire trial: 0.88 ± 0.14 ; familiar box: 0.91 ± 0.09 ; 95th percentile of shuffled distribution: 0.20).

Behavioural procedures. The rats were kept at ~90% of their free-feeding body weight and food deprived 12–18 h before each training or recording session. During the 2–3 weeks between surgery and testing, the animals were trained to collect randomly scattered vanilla or chocolate biscuit crumbs in either of two sets of environments. Rats from ref. 12 were trained in 1.5 m \times 1.5 m or 2.2 m \times 2.2 m square black 0.5 m high boxes with black floor mats and black walls. Rats in ref. 2 were trained in 1.0 m \times 1.0 m square boxes or 2.0 m wide circular boxes, all black and all with 0.5 m high walls. Abundant visual cues were available external to each recording box (Extended Data Fig. 1). The 100 cm wide squares were presented in both novel and familiar recording rooms, using an ABA design (familiar—novel—familiar). Different rooms were used for the 1.5 m and 2.2 m boxes. In all rooms, the boxes were placed at the back end of the room, with the recording system and experimenter at the front end, near the door. In ref. 12, a straight separation wall (1.5 m box) or a separation curtain with an opening (2.2 m box), was placed between the box and the experimenter. No separation was used in the 1.0 m square or in the circle. Recordings in the largest box lasted 30 min or occasionally more; in the 1.5 m box, the minimum duration was 15 min. Recordings in the 1.0 m box lasted 10 min or more; in the circle, trials were 20 min. Floor mats were always washed between trials. Before and between trials, the rat rested in the flowerpot on the pedestal next to the recording box.

Spike sorting and cell classification. Spike sorting was performed offline using graphical cluster-cutting software (tint, Neil Burgess, for Axona data; MClust, A. D. Redish, for Neuralynx data). Clustering was performed manually in two-dimensional projections of the multidimensional parameter space (consisting of waveform amplitudes and waveform energies), using autocorrelation and cross-correlation functions as additional separation tools and separation criteria. To ensure cells were not included in more than one data set, cells were compared across successive recording days. If two cells with similar spike clusters on the same tetrode had indistinguishable grid fields on two successive sessions, only one of them (the cell associated with the best coverage and the best signal-to-noise ratio) was included in further analysis. This approach drastically reduced the data set in some animals as large numbers of grids were recorded on multiple occasions (in one animal, 165/341 cells were discarded because of identified repeats).

Grid modules were defined by a *k*-means clustering algorithm as reported previously¹². The analysis was based on multidimensional data from each cell, consisting of grid spacing and grid orientation, as well as the eccentricity of the grid pattern. The *k* value was determined as the number of peaks detected in each animal's kernel smoothed density estimate of log grid spacing and grid orientation.

Rate maps. Position estimates were based on tracking of the LEDs on the head stage connected to the drive. All data were speed filtered; only epochs with instantaneous running speeds of 5 cm s⁻¹ or more were included.

To characterize firing fields, the position data were sorted into 3 cm \times 3 cm bins. The path was smoothed with a 21-sample boxcar window filter (400 ms; 10 samples on each side). Firing rate distributions were then determined by counting the number of spikes in each bin as well as the time spent per bin. Maps for number of spikes and time were smoothed individually using a boxcar average over the surrounding 5 \times 5 bins. Weights were distributed as follows:

```
box = {0.0025 0.0125 0.0200 0.0125 0.0025;...
0.0125 0.0625 0.1000 0.0625 0.0125;...
0.0200 0.1000 0.1600 0.1000 0.0200;...
0.0125 0.0625 0.1000 0.0625 0.0125;...
0.0025 0.0125 0.0200 0.0125 0.0025}
```

Identification of grid cells. The structure of the rate maps was evaluated for all cells with more than 100 spikes on the baseline session by calculating the spatial autocorrelation for each smoothed rate map³. The degree of spatial periodicity ('gridness' or 'grid scores') was determined by calculating the rotational symmetry of the cell's spatial autocorrelogram, as described previously¹². A cell was defined as a grid cell if its grid score exceeded a chance level determined by repeated shuffling of the experimental data (100 permutations per cell). For each permutation, the entire sequence of spikes fired by the cell was time-shifted along the animal's path by a random interval between on one side 20 s and on the other side 20 s less than the length of the session, with the end of the session wrapped to the beginning. A rate map and a spatial autocorrelation map were then constructed, and a grid score was calculated for each permutation. If the grid score from the recorded data was larger than the 95th percentile of grid scores in the distribution from the shuffled data, the cell was defined as a grid cell.

Definition of grid orientation. For fine-grained analysis of the geometric features of the grid cells, we analysed grid orientation in the following manner. From the spatial autocorrelograms, we defined individual fields as neighbouring bins above a correlation criterion (0–0.5, depending on scale). We then computed the centre of

mass for each of the six fields closest to the autocorrelogram centre field ('the inner ring'). As spatial autocorrelograms are symmetric, we defined 3 axes for further analysis. We first determined which inner field centre in the inner ring was closest in absolute angular distance to the nearest wall and labelled this axis 1. The two nearest axes were labelled axis 2 (most positive) and axis 3 (most negative).

Estimate of grid orientation and rotational offset. The minimal offset of any grid axes to any of the four walls in the box was determined as follows. From the matrix of grid orientation values across all cells, X (size: $n \times 3$), we identified the absolute angle to the nearest wall (in degrees):

$$A = 45 - \text{abs}([X \bmod 90] - 45)$$

We then sorted the grid axes according to A per cell and defined A_{\min} to be the smallest offset across axes per cell. By referring back to the original grid orientation for the sorted axes values in A , we could map each cell's anchoring (A_{\min}) to a particular wall pair (original axis closest to 0° : east–west axis, closest to $\pm 30^\circ$: north–south axis). Rayleigh test for circular uniformity are referred to as 'Rayleigh test' in the main text and figure legends.

Kernel smoothed density estimate of grid orientation. Kernel smoothed density estimates of grid orientation were computed by considering 100 points equally spaced that cover the range of data in x . Given samples x_1, \dots, x_n

$$f(x) = \frac{1}{n} \sum_{i=1}^n K_\sigma(x - x_i)$$

where K is the Gaussian kernel

$$K = e^{-\frac{x^2}{2\sigma^2}}$$

with width σ .

Ellipse fitting. We fitted ellipses to the grid pattern using a direct least squares fitting approach as described in ref. 12. For each cell, an ellipse was fitted to the x and y coordinates of the six innermost detected field centres of the autocorrelogram. Briefly, a general conic is represented by an implicit second order polynomial:

$$F(A, X) = AX = ax^2 + bxy + cy^2 + dx + ey + f = 0$$

where $X = [x^2 \ xy \ y^2 \ x \ y \ 1]^T$ and $A = [a \ b \ c \ d \ e \ f]^T$

The fitting of a general conic can be achieved by minimizing the sum of squared algebraic distances:

$$D_A(A) = \sum_{i=1}^n F(x_i)^2$$

For direct ellipse-specific fitting, the quadratic constraint $4ac - b^2 = 1$ was imposed and expressed in matrix form as $A^T C A = 1$ as

$$A^T \begin{bmatrix} 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} A = 1.$$

after which the constrained ellipse fitting problem was reduced to minimizing $E = DA^2$ subject to the constraint

$$A^T C A = 1$$

To determine the relationship between ellipticity and grid rotation, we estimated grid distortion by defining the lengths of the semi-major and semi-minor axes as α and β , respectively. To quantify the amount of grid pattern distortion, we used two ellipse measures. From the ellipse fit, we determined ellipse orientation or ellipse tilt (orientation of semi-major axis), ellipticity, and eccentricity. Ellipse eccentricity was calculated as:

$$\epsilon = \sqrt{1 - \left(\frac{\beta}{\alpha}\right)^2}$$

where α is the ellipse semi-major axis and β the semi-minor axis of the ellipse.

Ellipse eccentricity ranges between 0 and 1 and describes how much a conic section deviates from perfect circularity (where 0 represents a perfect circle and 1 the limit at which the ellipse becomes parabolic).

We also used a simpler measure with more intuitive properties in terms of translating ellipse elongation to grid pattern distortion, defined as ellipticity

$$\epsilon = \frac{\alpha}{\beta}$$

where α is the semi-major and β the semi-minor ellipse axes. Ellipticity ranges from 1 to ∞ where 1 represents a perfect circle and ∞ the limit at which the ellipse becomes parabolic. Ellipticity describes elongation of the ellipse (length of semi-major axis compared to the semi-minor axis). Ellipse orientation was expressed as the angle (between -90 and 90 degrees) of the ellipse semi-major axis to zero. Grid cells had an average ellipticity of 1.17 ± 0.004 and eccentricity of 0.48 ± 0.005 (mean \pm s.e.m.).

Deformation by shearing. Shearing was performed, for each cell, along shear axes parallel to the two box cardinal axes. The grid pattern used in these analyses was defined by the 3 grid axes detected from the innermost hexagon of field centres in the spatial autocorrelogram for each cell. The shear transform was defined for two dimensions by:

$$f(x, y) = \begin{bmatrix} 1 & \gamma_1 \\ \gamma_2 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

where γ_1 is the shear parameter along the y axis and γ_2 along the x -axis and x and y are row vectors of initial coordinates of points in the plane. For each cell, and shear axis, shearing was performed across a range of shear parameters γ (-0.55 to 0.55 in 0.001 steps). We then calculated the grid ellipticity resulting from each transform. The shear that produced the smallest ellipticity value was defined as the optimal shear parameter. We then calculated updated grid axes from this optimal shear in each direction.

Simulations. To simulate grid maps for square environments, a probability density function was generated for each cell in which Gaussian modes were distributed according to an equilateral grid pattern. Each mode was described by a two-dimensional Gaussian function:

$$f(x, y) = A \exp \left(- \left(\frac{(x - x_0)^2}{2\sigma_x^2} \right) - \left(\frac{(y - y_0)^2}{2\sigma_y^2} \right) \right)$$

where A is the amplitude, x_0 and y_0 are the coordinates for the peak location, and σ_x^2 and σ_y^2 the Gaussian variance in the two dimensions. We set $\sigma_y^2 = \sigma_x^2$. σ^2 was identical for each mode and determined by the inter-peak distance (inter-peak distance/5)². Location of nodes were determined from:

$$z = \phi e^{i\theta} = \phi (\cos \theta + i \sin \theta)$$

where z is complex such that peak locations in the x, y plane are given by $x = \text{real}(z)$ and $y = \text{imag}(z)$, and

$$\phi = \xi \sqrt{(a - c)^2 + ac}$$

$$\theta = \psi + \text{atan} \left(\frac{\sqrt{3}c}{2a - c} \right) + \frac{\pi(b - c)}{3a}$$

$$c = b \bmod a$$

where ξ is grid spacing, ψ grid orientation in radians, a the index for identity of the hexagonal ring of nodes and b the index for the identity of the node on that ring. To simulate alignment of the grid to a box axis, $\psi = 0$ or $\pi/6$. To shear the grid pattern, the shearing transform was applied to the real and imaginary components of points in z (above).

Interactive shearing patterns. We simulated the effect of multiple interactive shear forces on grid geometry. Simulated grids were generated as above. Grid spacing, ξ , was set to 0.45, and box frame as $[-1 \ -1 \ 1 \ 1 \ -1; -1 \ 1 \ 1 \ -1 \ -1]$, defining a square box with sides = 2. Only fields that fell within this box were used. A simple shear transform along the y axis involves adding the x component multiplied by the shear parameter γ to the y component of every ordered pair in $[x, y]$.

$$f(x, y) = \begin{bmatrix} (x - 1) \\ (y - 1) + \gamma x \end{bmatrix}$$

The $(x - 1)$ and $(y - 1)$ terms moves the shear origin (anchoring point) to the $[-1, -1]$ corner of the box.

To simulate shearing from two orthogonal walls we set

$$f(x, y) = \begin{bmatrix} (x - 1) + \gamma_1 y \\ (y - 1) + \gamma_2 x \end{bmatrix}$$

where γ_1 was set to $\tan(\pi/24)$ as the original shear parameter such that an offset of 7.5° was induced along one axis. We then systematically varied γ_2 between $[-0.35, 0.35]$ in 0.025 steps. Rate maps were generated using the two-dimensional kernel smoothed density estimator described above (Gaussian kernel width: 0.1) within the box frame (62,500 evenly spaced evaluation points) and spatial autocorrelograms were generated from these. We then determined grid features from these autocorrelograms as described above.

To simulate symmetric shearing from two opposite (east–west) walls, we set

$$f(x, y) = \begin{bmatrix} x(1 + \gamma) \\ y(1 + \gamma x) \end{bmatrix}$$

and calculated grid features from the resulting maps. To generate grids that were fractured along the diagonal, we first generated a sheared grid according to:

$$f(x, y) = \begin{bmatrix} (x - 1) \\ (y - 1) + \gamma x \end{bmatrix}$$

We then discarded all the points $x + y < 0$ (below the diagonal of the box), and added a copy of the remaining points mirrored around that diagonal

$$f(x_s, y_s) = \begin{bmatrix} x_s & -x_s \\ y_s & -y_s \end{bmatrix}$$

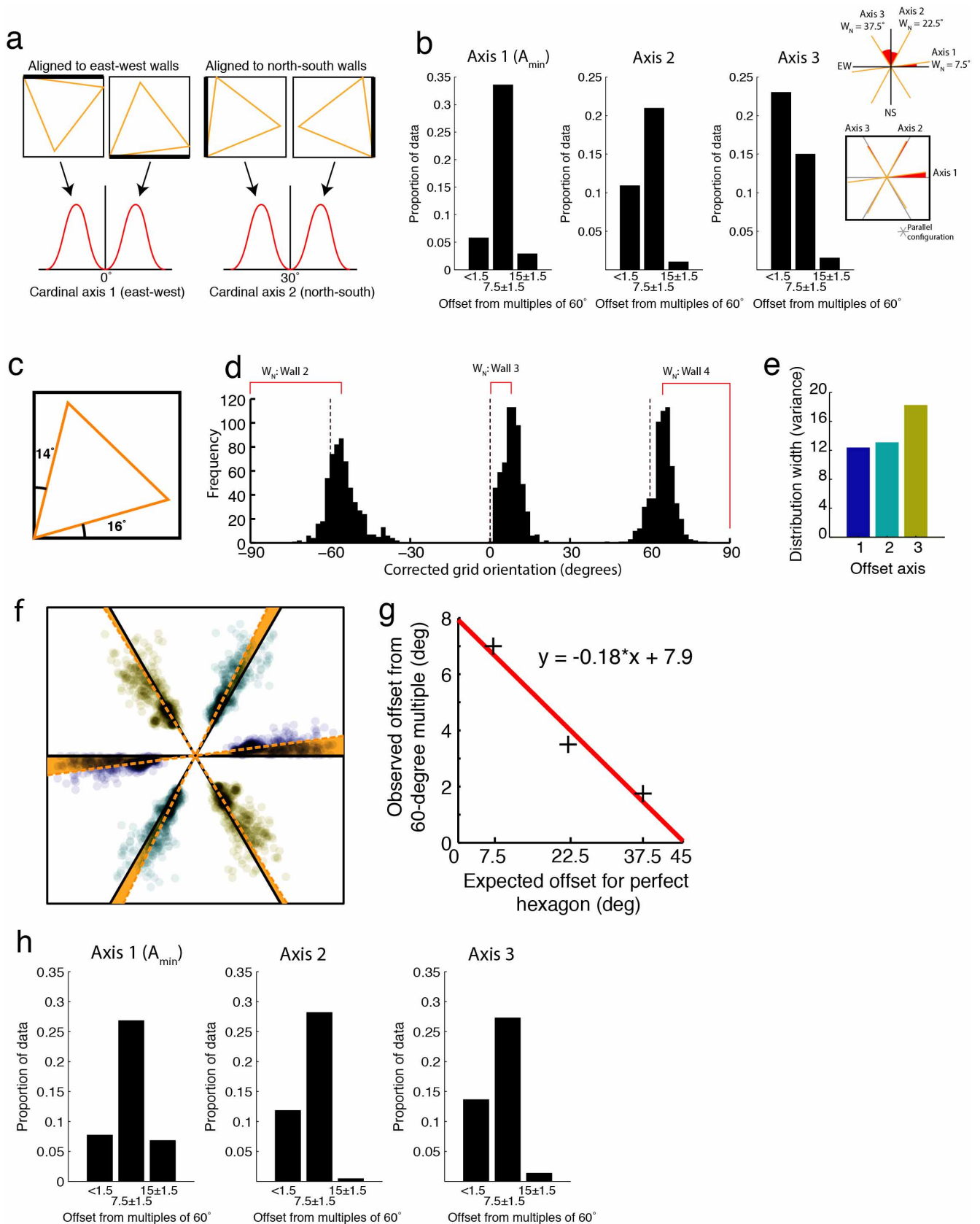
where x_s and y_s represent the sheared points from the original matrix $[x, y]$, whose sum was > 0 .

Local anchoring. To generate subcompartment maps, the firing rate map of each grid cell was divided into 4 or 9 equal quadrants. This analysis was done on 5 animals (3 from the 1.5 m box and 2 from the 2.2 m box; 17 modules in total) in which 4 modules were recorded. Then autocorrelograms for these were made, and grid features determined according to methods described above. Only modules that reliably yielded 6 detectable fields were used for analyses.

Histology and reconstruction of tetrode placement. Histology and reconstruction of electrode positions was performed as described previously, using flatmaps for two-dimensional location of electrode locations within the MEC and adjacent parasubiculum.

Statistical tests. Statistical tests were two-sided unless otherwise specified. No statistical methods were used to predetermine sample size.

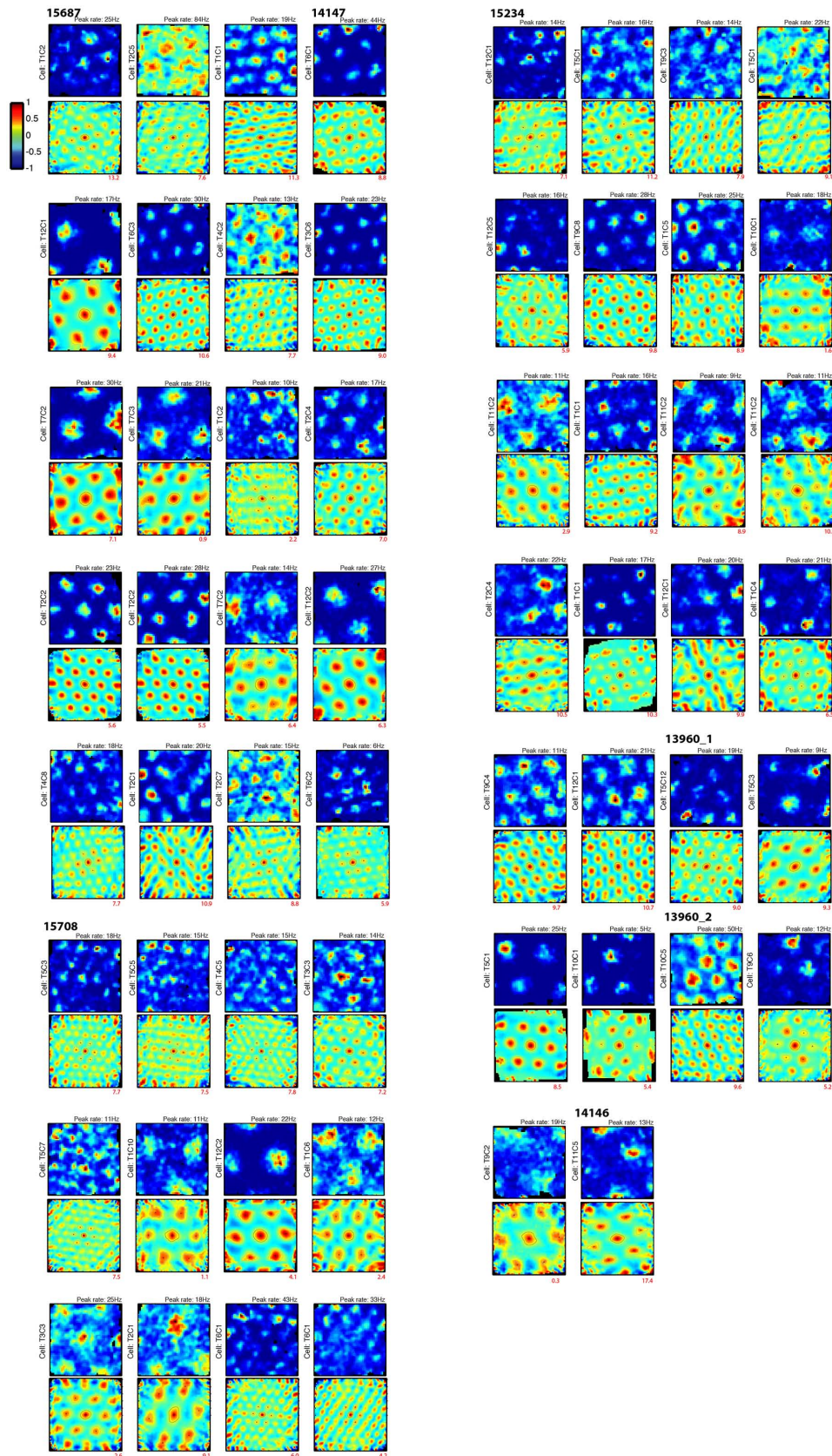
Code availability. Code for shearing transformations can be obtained from the authors.



Extended Data Figure 1 | Alignment of individual grid axes to box

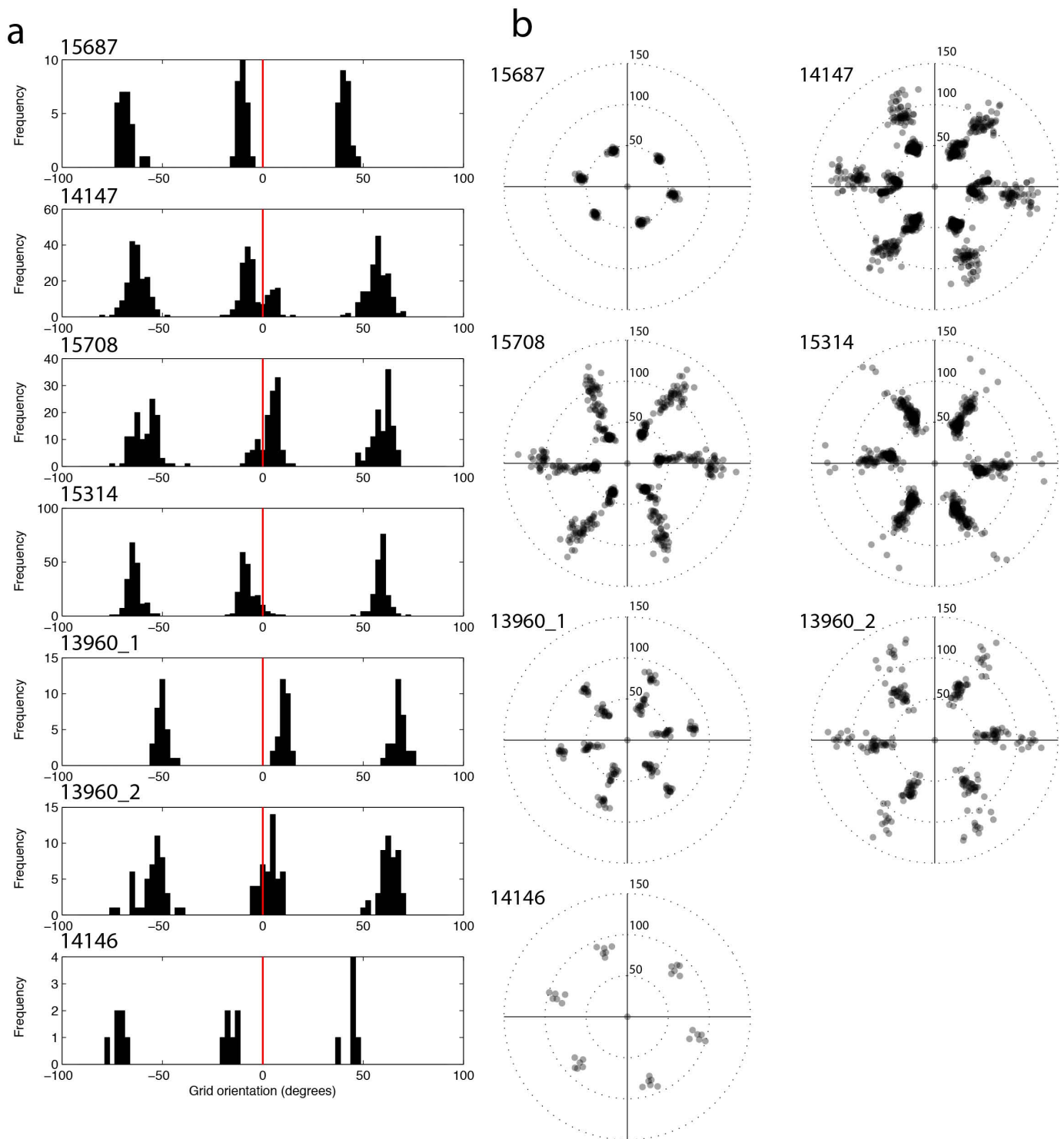
geometry. **a**, Interpretation of alignment configurations. Bimodal distributions (red curves) around the two cardinal axes (bottom) can be mapped as unique alignment solutions to the cardinal axes x and y of a square box (top row). Orange triangles show how the basic unit of the grid pattern can be uniquely aligned to either cardinal axis (thick solid black lines). **b**, Distribution of angular offset for all 3 grid axes of all cells recorded in the 1.5 m box. Amount of data are shown for the two symmetric alignment configurations 0° and 15° (symmetry between box geometry and grid) and for the offset where overlap between box and grid axes is minimized (7.5°). For each cell, grid axes were sorted according to angular distance from the nearest wall (labelled axis 1–3 with increasing wall-distance (top inset)). The absolute offset from the nearest multiple of 60° was calculated for each axis as a measure of the deviation from a perfectly hexagonal grid parallel to the east–west axis (bottom inset). Frequency plots show these distributions with 3° wide bins (x axes). The offset for grid axis 1 (A_{\min}) was consistently larger than for axes 2 and 3. For axis 1, 33.6% of the absolute values for A_{\min} fell within $7.5 \pm 1.5^\circ$, 5.8% within $0\text{--}1.5^\circ$, and only 2.9% within $15 \pm 1.5^\circ$. **c**, Schematic illustrating that the farthest away any axis of a perfect grid can be from any wall in the environment is 15° because of the symmetries inherent in the grid pattern and the square box. Because the offset distribution is constrained to $[0, 15]^\circ$, tests against uniformity were performed by multiplying the offsets by 24 to achieve a 360° range. **d**, Frequency plot of grid orientation for axes 1–3 for all cells in the 1.5 m environment after correcting for alignment direction (all cells with $A_{\min} < 0$ were multiplied by -1 across the alignment axis (reflected)). Distance to W_N is

indicated by red lines for each axis. **e**, Frequency plot showing distribution width, or variance, for grid axes as a function of distance from W_N in the 1.5 m environment. Distributions of the 3 grid axes were sorted by distance to W_N (as in **b**) and are shown in distinct colours (as in Fig. 1g). **f**, Polar scatterplot (same data as in Fig. 1b) showing axis-specific offset from W_N or 60° multiple in the 1.5 m box. All grids with $A_{\min} < 0$ have been reflected around the A_{\min} wall axis in order to re-align axes to one absolute offset solution. Colours as in Fig. 1g. Orange dashed lines show peak angular offset from 60° multiples (black lines) of parallel wall alignment (0°) for each axis. Orange wedges highlight the angle between each axis and its nearest 60° multiple of 0° . **g**, Relationship between expected angular offset from W_N for a perfectly hexagonal grid with 7.5° angular offset on axis 1, and observed offset from 60° multiple for each grid axis in the 1.5 m box (mean of all cells). Least square regression line is indicated. Note that the observed offset decreases as the offset from W_N approaches its maximum at the diagonal (45°). **h**, Frequency plots showing only slightly larger angular offset along axis 1 (the axis nearest W_N) than axes 2 and 3 in the 2.2 m environment. For axis 1, a total of 7.7% of the cells from the 2.2 m box had minimal angular offsets within $0\text{--}1.5^\circ$ of the nearest wall, 26.8% had offsets of $7.5 \pm 1.5^\circ$, and only 6.8% were rotated $15 \pm 1.5^\circ$. In 203/220 cells, grid orientation could be referred back to either of 3 distinct anchoring solutions. The mean offsets from the walls were similar, all centering around 7.5° (mean \pm s.e.m. A_{\min} : north–south wall negative offset: ($n = 108$), $7.54 \pm 0.03^\circ$; positive offset ($n = 58$), $6.35 \pm 0.07^\circ$; east–west wall negative offset ($n = 37$), $8.16 \pm 0.12^\circ$). Offsets were largely unimodal in the 2.2 m box (Fig. 3c), in contrast to the 1.5 m environment (Fig. 1c).



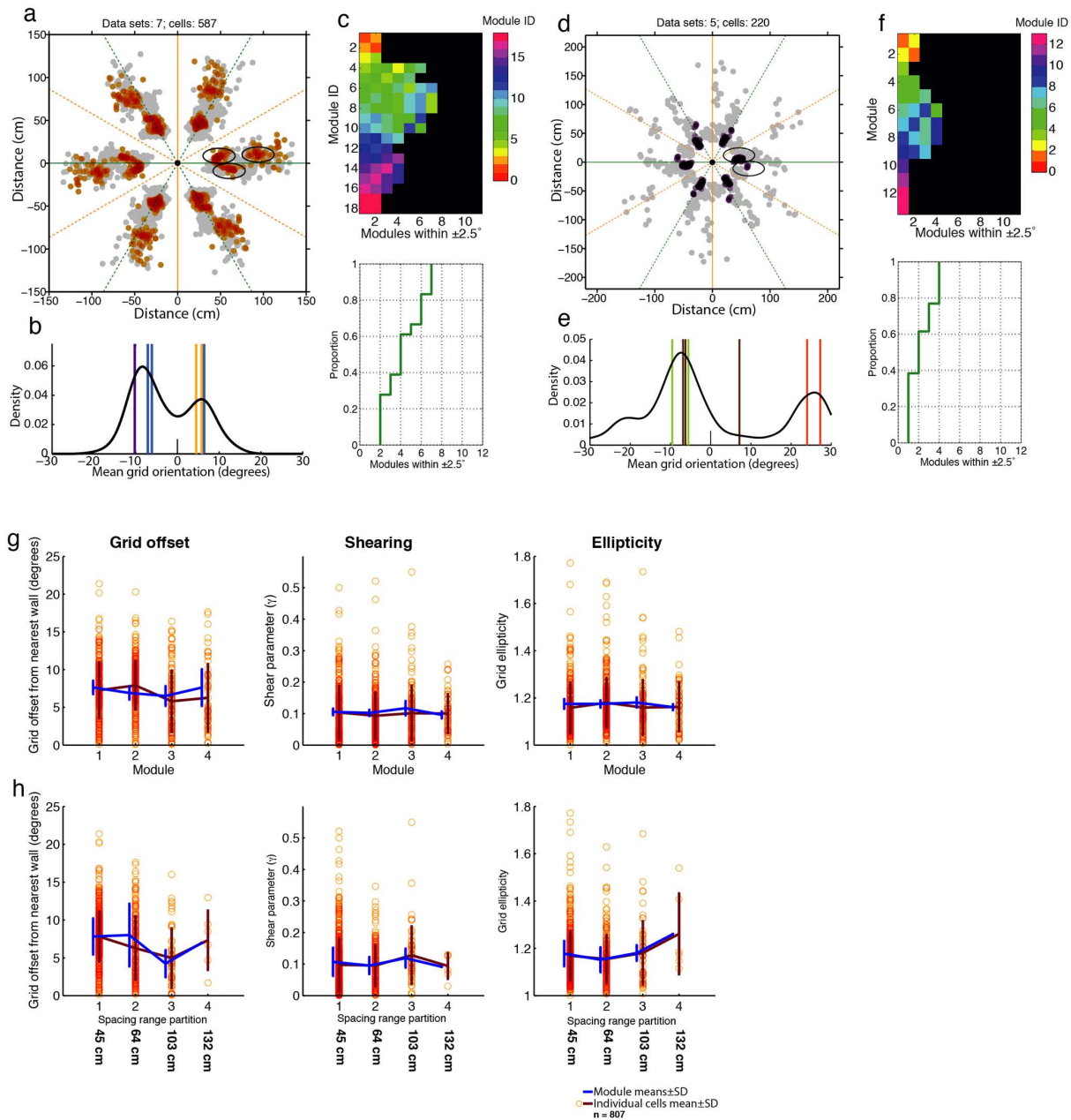
Extended Data Figure 2 | Rate map and spatial autocorrelation diagram for representative cells from all animals. The 587 cells from the 1.5 m environment were ranked according to time of recording for each animal and every tenth cell on the list was then selected. Cells are sorted by animals. For each cell, the rate map is shown at the top and the autocorrelation at the bottom. Colour scale, 0 Hz to peak rate for rate maps; $[-1,1]$ for

autocorrelation maps. Peak rate is indicated above the rate map and gridness below the autocorrelation diagram. Correlation is indicated by scale bar. Peak rate is indicated at the bottom of each rate map; angular offset at the top right of each autocorrelation diagram. Grid orientation and grid ellipticity were determined from the innermost hexagon of vertices in the autocorrelation diagram, as in Fig. 1a.



Extended Data Figure 3 | Distribution of grid orientation (A_{\min}) across individual animals in the 1.5 m environment. **a**, Frequency distributions showing clustering of grid orientation around each of the 3 axes defined by 60° multiples of the east–west box axis (number of cells as a function of orientation). Each row shows one animal; animal number is indicated at the top left. Rat 13960 had two distinct data sets (1 and 2, respectively). **b**, Polar scatterplots showing distribution of grid orientation and grid spacing for the entire sample of grid cells in each animal. Grid spacing is indicated on the radial

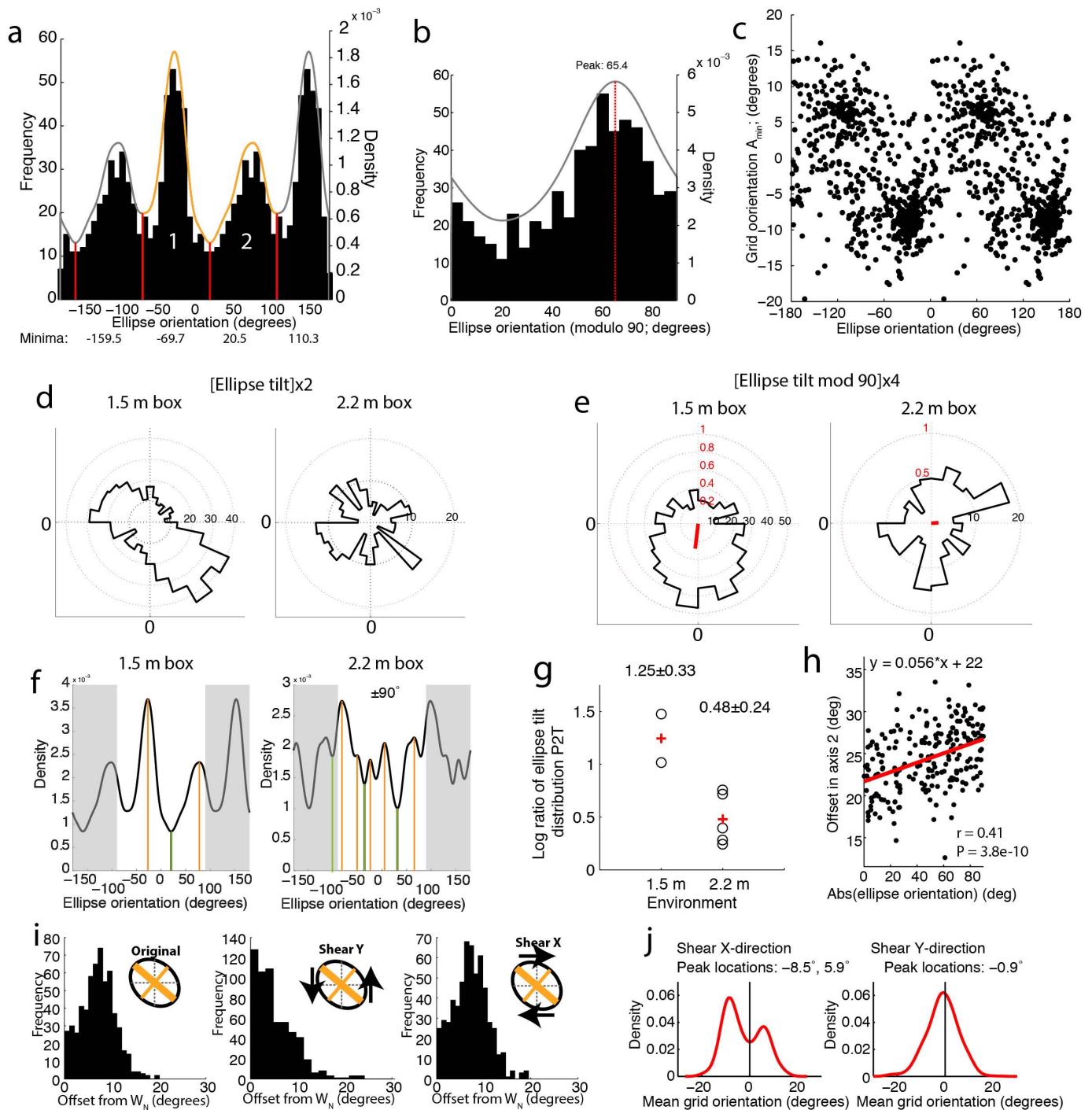
axis. For each cell, the location of the 6 innermost fields in the spatial autocorrelogram is shown, as in Fig. 1a, b. The cardinal axes of the recording environment are shown as black lines. Note similar offset from the east-west axis in at least 6 out of 7 data sets (all animals except 14146). The deviation from the parallel configuration (0°) was consistent across animals. Excluding cells with fewer than 6 grid fields did not cause any major change of the angular offset (mean \pm s.e.m. of remaining cells: $7.5 \pm 3.3^\circ$).



Extended Data Figure 4 | Modular organization of grid alignment.

a–c, Data from 1.5 m environment. **a**, Polar scatterplot of all cells recorded in the 1.5 m box (grey), as in Fig. 1b. Superimposed are all points from one animal (orange points with multiplicative colour). Green ellipses indicate 3 different modules. Different modules assumed either of the two alignment configurations around the east–west box axis, that is, distributions for individual modules were always unimodal. **b**, Kernel smoothed density curve showing frequency of observations across values of A_{\min} (all cells; kernel width: 2.5°). Coloured lines show means of individual grid modules. Colours correspond to different animals. Module means showed the same distribution trend as the pooled data. The average of the smallest angular offset (A_{\min}) with grid modules as the unit of analysis was similar to the average of the pooled data (mean \pm s.e.m.: $6.3 \pm 0.2^\circ$). Both within and across animals, the distribution was bimodal. **c**, To visualize cross-animal clustering of grid orientation, for each grid module we counted the number of other modules with mean grid orientations that distributed within a narrow range of the mean orientation of the reference module ($\pm 2.5^\circ$). Left plot shows for each module (in individual rows) which other module means fell within $\pm 2.5^\circ$. Modules were sorted according to mean grid orientation and assigned a rank value (module ID). Each module has a unique colour based on module ID (colour-bar). The grid orientation of a module was typically shared by several other modules (4.2 ± 1.9 module means within $\pm 2.5^\circ$), as indicated by many long

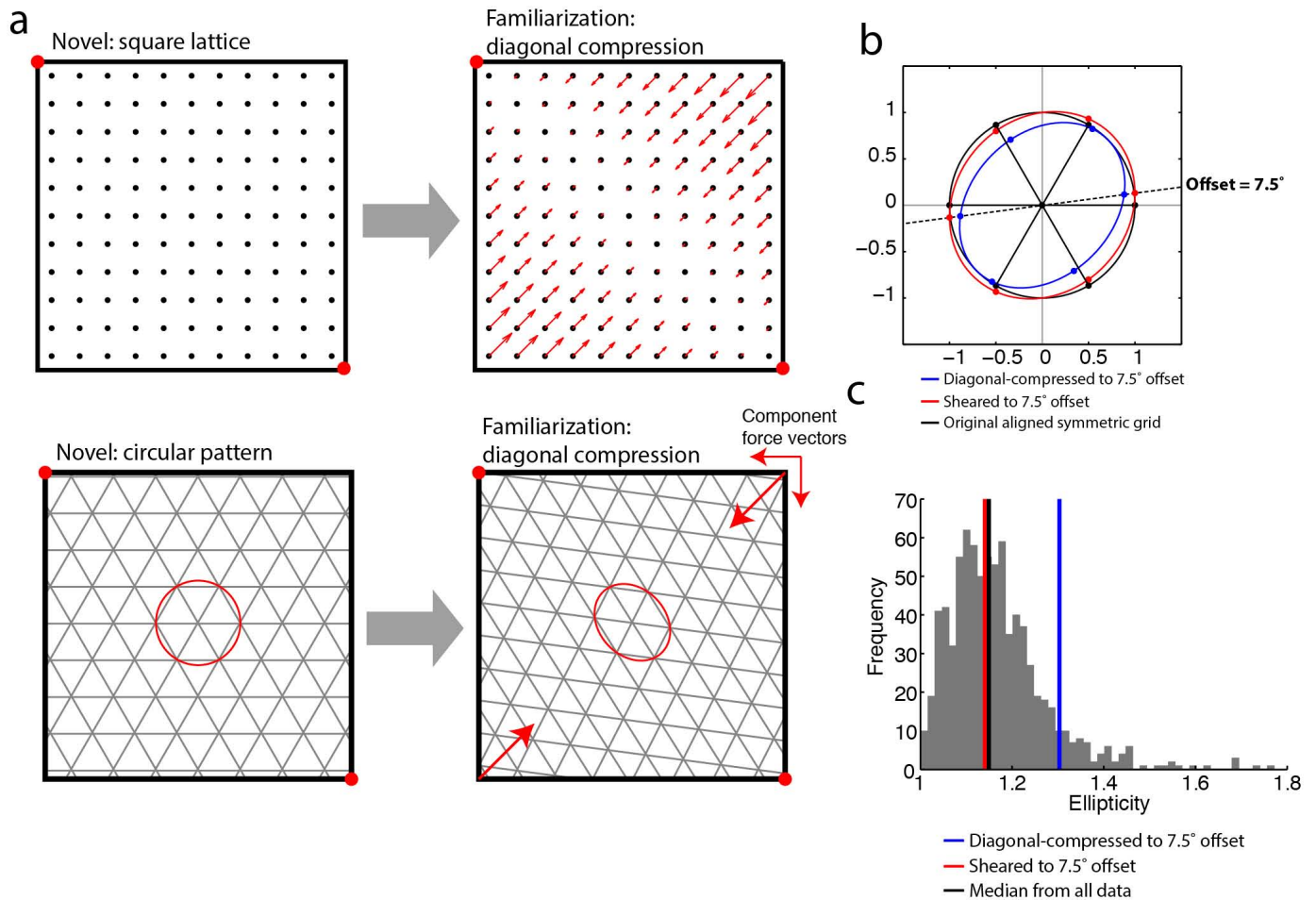
rows as well as consistent clustering of colour across rows). Bottom plot shows cumulative distribution of the same data. Taken together, the data suggest that the number of anchoring solutions is limited and that solutions are reused across modules. **d–f**, Data from 2.2 m environment. **d**, Polar scatterplot with an individual animal highlighted in purple. **e**, Distribution of mean values for grid orientation across modules, as in **b**. The average minimal orientation offset across module means in the 2.2 m box was $7.2 \pm 0.3^\circ$ (mean \pm s.e.m.). **f**, Clustering of mean grid orientation across modules. Notation as in **c**. **g, h**, Distortion and rotation of grid pattern are independent of grid spacing. **g**, Scatterplots showing relationship between module identity and grid offset (left), optimal shearing parameter (middle), and grid ellipticity (right). Each circle corresponds to one cell. Dark red lines show means \pm s.d. for successive modules based on pooled data of individual cells. Blue lines show means \pm s.d. for mean values of each module (slightly offset in x direction for clarity). Data from 1.5 m and 2.2 m environments are combined. Note lack of effect of module identity on angular offset, shearing parameter or ellipticity. **h**, Same plots as in **g**, but binned by grid spacing. Spacing range was partitioned into 4 equal parts. The values underneath each panel show the centre value of each partition. The absolute offset decreased only minimally with grid spacing (repeated-measures Kruskal–Wallis test with 4 bins of grid spacing as the within-subjects factor: $H = 0.8$, d.f. = 3, $P = 0.85$; mean offset for modules M1–M2: 7.3° ; mean offset for M3–M4: 6.9°).



Extended Data Figure 5 | Grid deformation is constrained by environmental geometry.

a–c, Data from 1.5 m environment. **a**, Frequency plot showing orientation (tilt) of the semi-major axis of an ellipse fit to the inner hexagon of firing vertices in each cell's spatial autocorrelogram (all 587 cells that were recorded in the 1.5 m box). For clarity, ellipse values are shown over the entire 360° range (each cell thus contributes 2 data points as unique ellipse orientation values distribute within a 180° range). The superimposed curve shows a kernel smoothed density estimate of the same data (Gaussian kernel width: 7.5°). The distribution of ellipse orientation was highly non-uniform both within and between animals (Rayleigh test in the entire cell sample: $Z = 12.8$, $P = 2.5 \times 10^{-6}$ for (ellipse orientation distribution) $\times 2$). Two unique modes within $[-90, 90]^\circ$ were identifiable. Modes were defined as all the values that fell between the two neighbouring troughs (red lines that mark the boundary of each mode-distribution, labelled 1 and 2, respectively). Cells within modules shared ellipse orientation. On average, for one mode, the long axis of the ellipse was offset by $27.7 \pm 1.1^\circ$ from the east–west axis of the recording box; for the other mode, the offset was $17.9 \pm 1.4^\circ$ from the north–south axis (means \pm s.e.m.). To determine if the spread of ellipse orientation was smaller than expected by chance, we drew, for each module, m random ellipse orientation values from the pool of all cells ($n = 587$), and determined the circular spread (standard deviation) of that distribution. m was the number of cells in the module. The procedure was repeated 300 times per module to allow calculation of z scores. The circular spread of ellipse orientation within modules was significantly lower than expected by chance (mean \pm s.e.m. $z = 4.53 \pm 1.33$; $t(17) = 3.37$ $P = 0.006$, one sample t -test). **b**, Frequency of the same data as in **a**, but modulo 90° to control for differential ellipse alignment (modulo 90° brings out trends common to each of the 4 walls in the box). A single peak was detectable at 65.4° (red line). **c**, Minimal wall offset (A_{\min}) as a function of ellipse orientation for each cell. Values are shown for an ellipse range of 360°, such that each cell contributes 2 points. Each mode of the distribution of ellipse orientations consisted primarily of cells from one of the two configurations for grid orientation (positive or negative offset). **d–g**, Comparison of deformation in the 1.5 and 2.2 m boxes. **d**, Circular histograms showing ellipse orientation for all cells recorded in the 1.5 m box (left) and the 2.2 m box (right). Ellipse orientations were multiplied by 2 to achieve a 360° range. Although there was clear clustering around two modes in the smaller box, no clustering was apparent in the data from the large box.

e, Same as in **d**, but for ellipse orientation modulo 90° and multiplied by 4 to achieve a 360° range (left, 1.5 m box; right, 2.2 m box). The modulo operation brings out distribution trends that are similar for the 4 walls in the box. No such trends were apparent for the 2.2 m environment. Mean vector length for the distributions is shown in the centre of each plot (red solid line and axis). **f**, Kernel smoothed density curves from ellipse orientation in all cells recorded in the 1.5 m box (left) and the 2.2 m box (right). We identified all peaks (orange lines) within $[-90, 90]^\circ$ (light areas), and then for each peak calculated the log ratio between the peak and the smaller of its two neighbouring troughs (green lines). While only 2 peaks were identified for ellipse orientation in the 1.5 m box (**a**), 5 peaks were present in the 2.2 m box. **g**, Log peak-trough ratios from the values calculated in **f** (means \pm s.e.m.). Values were significantly lower in the 2.2 m environment than the 1.5 m environment (two sample t -test, $t(5) = 3.51$, $P = 0.02$). **h**, Scatterplot showing, for all cells in the 2.2 m environment, grid orientation from the grid axis second furthest away from the nearest wall (W_N) versus absolute ellipse orientation. There was a strong linear correlation between these parameters ($r = 0.41$, $P = 3.82 \times 10^{-10}$), as expected if interactions from a second shear axis had occurred (linear regression is shown as red line). **i, j**, Distribution of angular offset after shearing along each of the cardinal axes of the 1.5 m box. **i**, Frequency plots showing minimal absolute offset from multiples of 60° of the parallel (0°) configuration for all cells before shearing (left) and after shearing in the north–south (y) and east–west (x) directions (middle and right, respectively; insets). For each box axis, we used the shear transform that minimized ellipticity. Before shearing, the distribution showed a clear offset (left). After shearing in the y direction (north–south; middle panel), the angular offset was eliminated. Shearing along the x axis (east–west) did not reduce the offset (right), despite similar reduction of the ellipticity. Note that, in simulated grids, the shear parameter γ needed to rotate a perfectly hexagonal grid from 0° to 7.5° ($\gamma = \tan(\pi/24)$) induced ellipticity of 1.16, a value almost identical to the mean ellipticity observed in the data (1.17). **j**, Kernel smoothed density estimates of A_{\min} (minimal offset from any wall) for all cells (Gaussian kernel width: 2.5°) after shearing along both the x and y axis (left and right panel, respectively). There were two distinct modes after shearing in the x direction. The distribution for the y direction, in contrast, became unimodal and centred on the east–west box axis (0°), suggesting that the bimodal distribution of angular offset was caused by shearing in this direction.



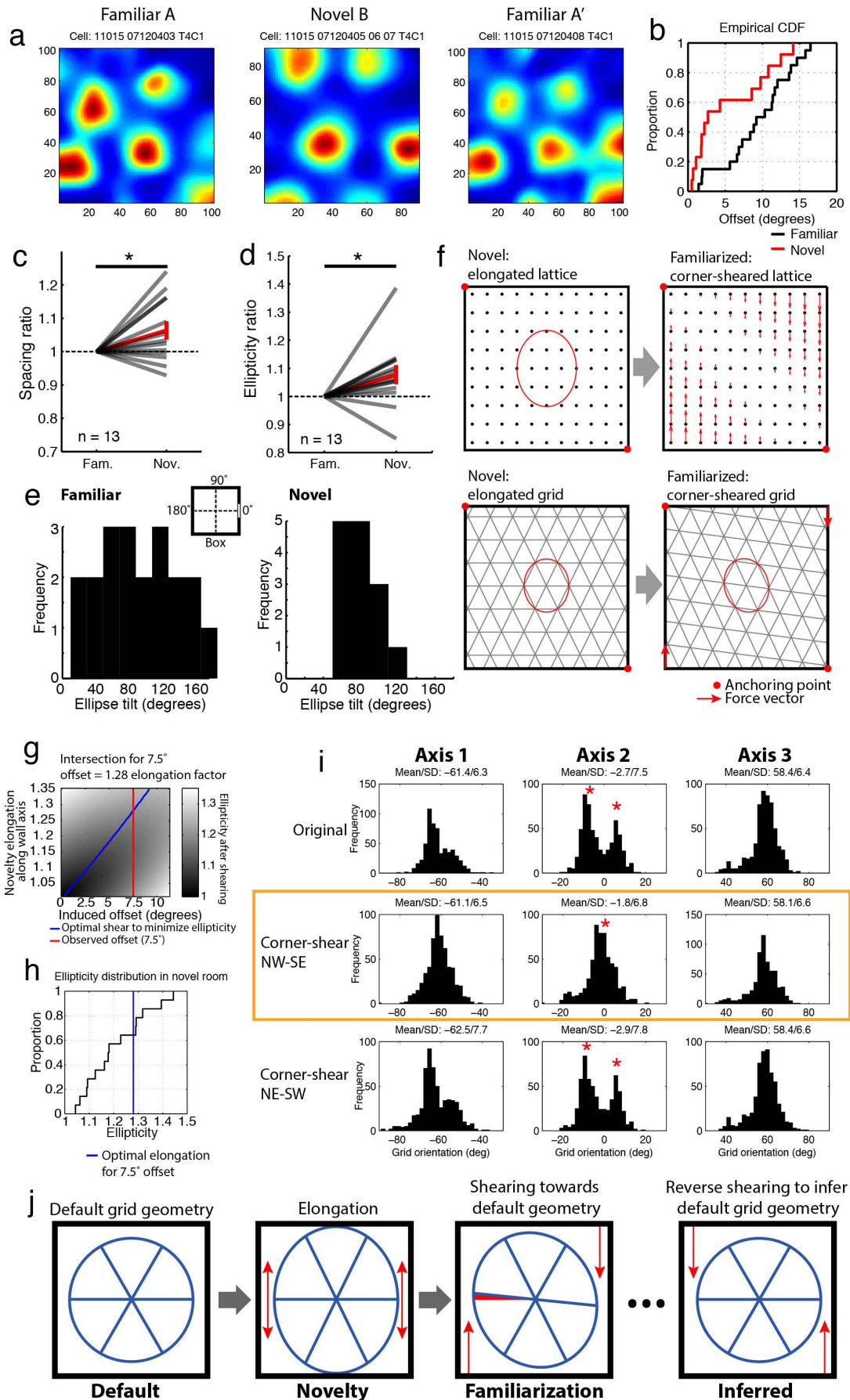
Extended Data Figure 6 | Observed grid deformation and rotational offset cannot be reproduced by mere diagonal compression, where forces from the corners move fields perpendicularly towards the diagonal. **a**, Diagonal compression model showing forces acting on a shrinking pattern that is anchored to opposite corners. Top left, square lattice (no elongation) representing starting point in novel environment. Top right, vector field (red arrows) of proposed forces resulting from pattern shrinkage given anchoring to opposite corners (red dots). Without elongation in novelty, pattern shrinkage during familiarization to the environment is hypothesized to be radially uniform. However, corner anchoring will stop shrinkage along the diagonal of the anchoring corners. This results in a compression of the pattern along the opposite diagonal. Bottom left, a simulated grid pattern that was uniformly enlarged (as indicated by the perfect circle fit) to mimic a novelty response. Upon familiarization to the environment (shrinkage, bottom right), and in the presence of anchoring to opposite corners, the pattern undergoes diagonal compression. The resulting pattern has an angular offset (here 7.5°) and is accompanied by pattern deformation. Force vectors are shown as red arrows inside the box, and component vectors are shown as red arrows outside the box. The diagonal compression transform is described by:

$$f(x,y) = \begin{bmatrix} 1+c & c \\ c & 1+c \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

where, in order to offset an aligned grid by 7.5°,

$$c = -\frac{1}{2} \left[1 - \tan\left(\frac{5\pi}{24}\right) \right]$$

b, Comparison of grid pattern characteristics after shearing and after diagonal compression, both aimed at achieving a 7.5° angular offset. Black lines and dots show the original symmetric and aligned grid pattern. Red dots and ellipse denote pattern deformation imparted by shearing, while blue dots and ellipse show deformation caused by diagonal compression. 7.5° offset is indicated by a dashed black line. Note the pronounced elliptical deformation of the diagonally compressed pattern compared to the sheared pattern. **c**, Histogram of ellipticity values observed in 1.5 m and 2.2 m recording environments. The median is indicated by a black line. Red and blue lines show ellipticity after 7.5° offset induced by shearing and diagonal compression, respectively. The ellipticity induced by shearing matches the data almost perfectly, while the value for diagonal compression is much higher than observed. Together, these analyses suggest that although diagonal compression can induce both angular offset and elliptic deformation, the relationship between these variables does not render this transformation a likely candidate for explaining the observed grid alignment. In contrast, the relationship between ellipticity and offset following shearing closely matches observation (Extended Data Fig. 7), suggesting that this is indeed the process that underlies the alignment process.



Extended Data Figure 7 | Experience-dependent changes in grid patterns can be explained by shearing of grids that are anchored to diagonally opposite corners.

To address effects of experience, grid orientation was compared at two stages of acquaintance with the environment, during the first trial of exploration in a novel square box in a novel test room and during exploration of a similar but highly familiar box in a familiar room. We used a previously published set of 20 grid cells from 5 rats that ran in 1 m wide square boxes with shape and colour similar to the 1.5 and 2.2 m boxes in the main study². 85% of the 20 grid cells had 6 fields or more. **a**, Rate maps for a representative grid cell recorded in a 1 m wide square box in a familiar room (left), a novel room (middle), and a second time in the familiar room (right). **b**, Cumulative distribution frequency plot showing distribution of orientation offsets in novel and familiar rooms for cells that were recorded in both environments. **c**, **d**, Ratio of grid spacing (**c**) and grid ellipticity (**d**) in the novel and the familiar environment (novel/familiar). Red line shows mean (\pm s.e.m.). * $P < 0.05$. Grid spacing and grid ellipticity increased from familiar to novel by factors of 1.06 ± 0.03 and 1.08 ± 0.03 , respectively (one-sample t -test of log ratios: $t(12) = 2.25$, $P < 0.05$ and $t(12) = 2.30$, $P < 0.05$; all cells recorded in both environments), consistent with previous observations of grid cells in novel environments¹⁸. **e**, Distribution of ellipse orientation in familiar and novel environments (left and right, respectively). Orientation is expressed in relation to the east–west wall (inset). Ellipse orientation was more sharply distributed in the novel environment, with a circular mean of 81.1° and s.d. of 16.7° , near orthogonal to the anchoring wall. In the familiar environment, the distribution was broader (circular s.d. of 37.4°), suggesting that the impact of the walls is relaxed with repeated experience. **f**, Analyses of simulated grid patterns showing that both de-elliptification and non-coaxial rotation can be induced also with elongated grid patterns as the starting point, if the grid is anchored to opposite corners of the recording box. Top left, square lattice anchored to opposite corners of the recording box. Top right, square lattice parallel to the east–west wall axis, elongated along the north–south wall axis. An ellipse is fitted to indicate degree of elongation (red). Top right, same lattice but allowed to relax towards a less elongated state. If the grid is anchored to opposing corners (red dots), a compression shear (corner shear) vector displacement field is generated (red arrows). Bottom left, simulated grid pattern, aligned and elongated in the same way as the lattice above. The elongation produces an ellipse (red) oriented along the north–south axis. No rotation is present at this initial stage. Bottom right, relaxation towards a less deformed state, with the grid anchored to opposing corners, produces a sheared pattern with an angular offset. Shearing force vectors are shown as red arrows in each of the non-anchoring corners. The transform is defined by:

$$f(x,y) = \begin{bmatrix} 1+\gamma_1 & \gamma_1 \\ \gamma_2 & 1+\gamma_2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

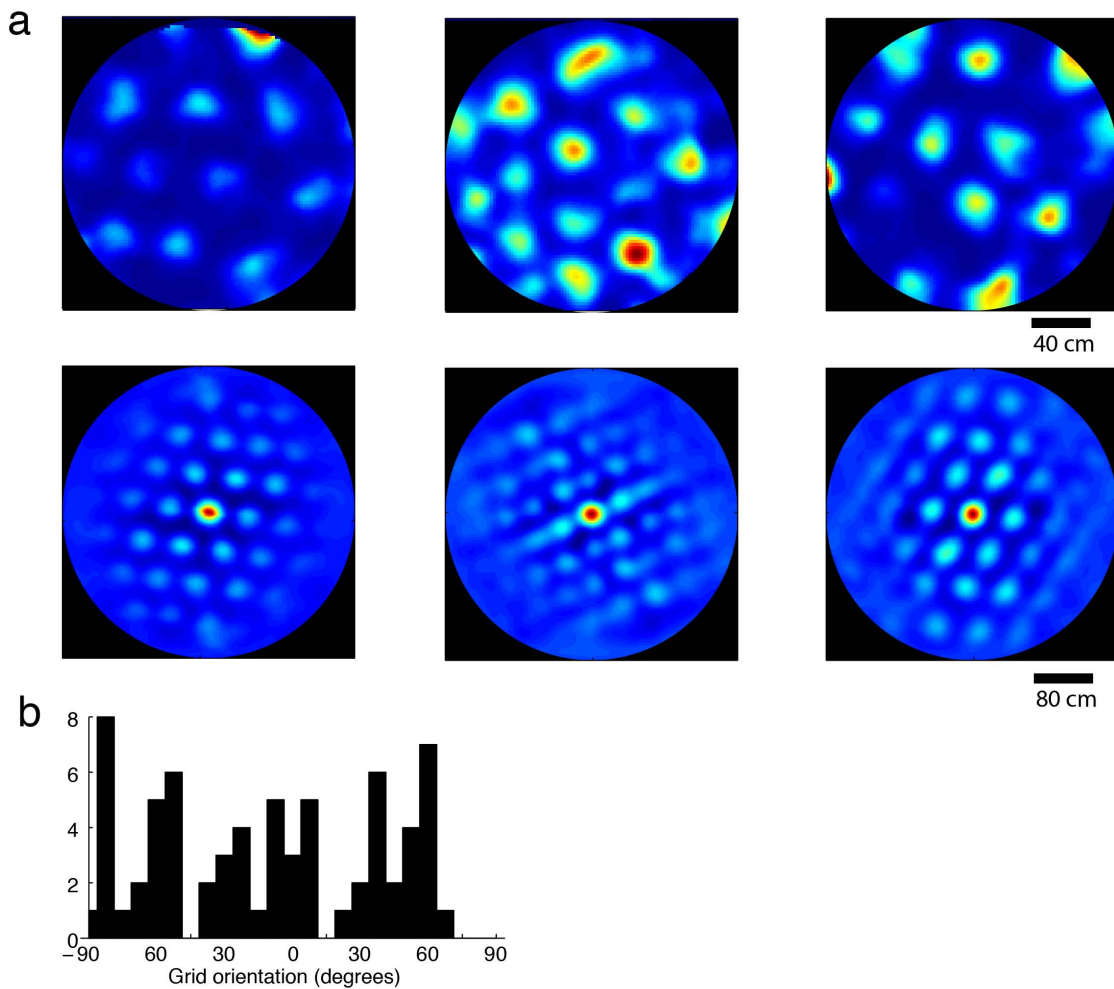
or

$$f(x,y) = \begin{bmatrix} 1+\gamma_1 & -\gamma_1 \\ -\gamma_2 & 1+\gamma_2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

depending on which corner pairs operate as anchors. So long as only one γ is non-zero, the transform produces shear-like displacement along one of the cardinal axes, but the shear axis is rotated 45° so as to leave the two anchor corners in place. Corner shearing minimized ellipticity and removed the angular offset of the pattern (see **i**). **g**, **h**, Analysis of relationship between ellipticity, angular offset and elongation level in novel environments.

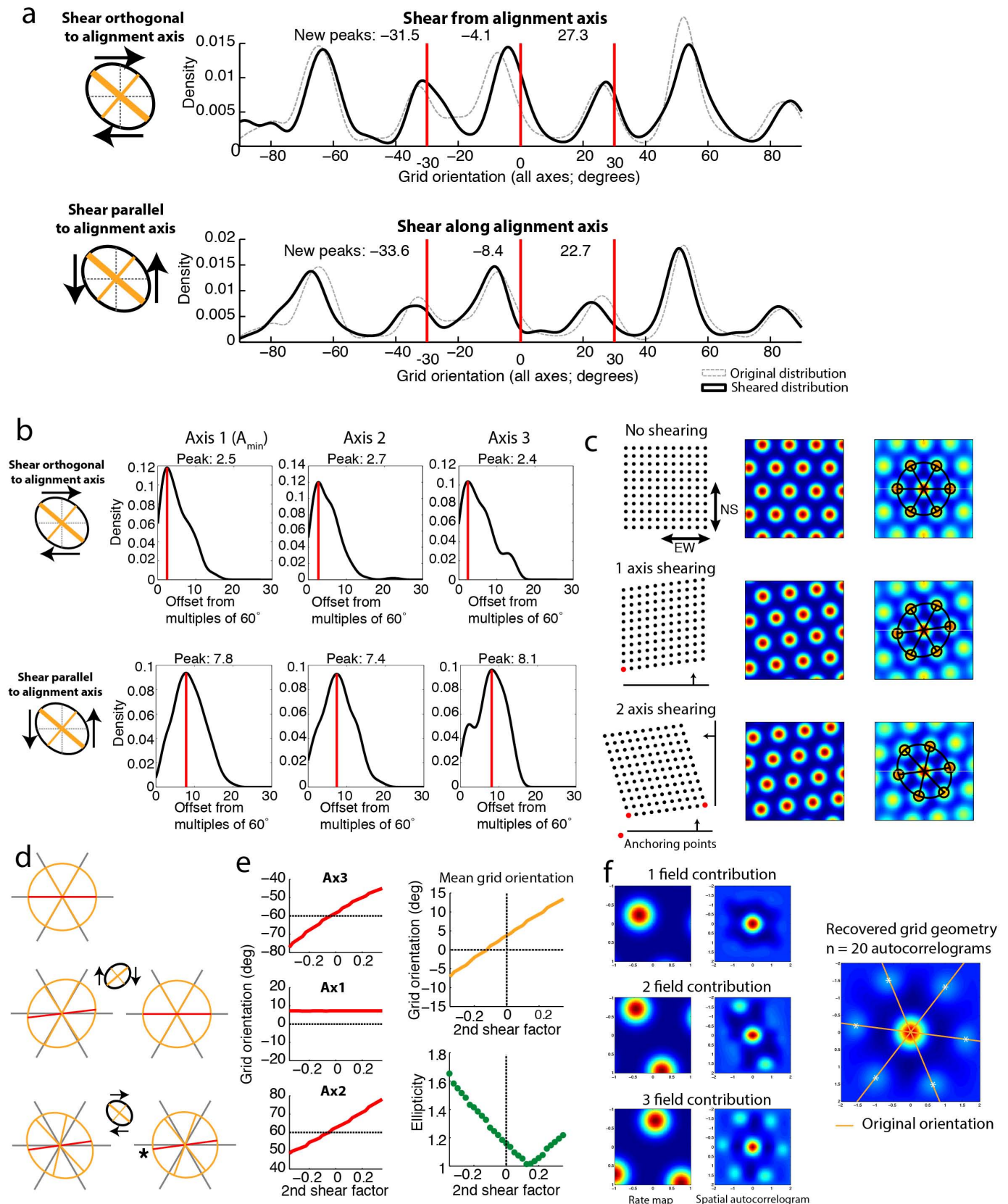
g, Optimal corner-shearing for minimizing ellipticity in an elongated grid. Ellipticity values are brightness-coded. Changes in grid orientation may take place after minimization of the elongation of the grid when the grid is anchored to opposite corners. For such changes to occur, shearing should minimize ellipticity given the initial elongated state. To test this idea, we systematically elongated simulated grid patterns along the wall axis perpendicular to the grid alignment axis (y axis in plot). For each elongation level, we next corner-sheared this pattern (with corners as anchoring points), through a range of shear parameters (x axis in plot, values correspond to the offsets associated with each shear parameter [offset = \tan^{-1} (shear parameter)]). For all elongation factors > 1 , shearing to a level > 0 produced less ellipticity than in the original pattern. This was not the case for the standard shear transform in which no corner anchoring occurs (not shown). Across elongation levels, we determined which shearing parameter minimized ellipticity (blue line). We next found the intersection between these optima and the shear parameter that produces a 7.5° offset (red line). The elongation factor at this intersection was 1.28. Thus, in order for a 7.5° offset to represent the optimal amount of shearing in simulated data with anchoring to diagonally opposite corners, the initial elongation factor of the grid pattern is 1.28, very close to the observed ellipticity in the novel environment (panel **h**). **h**, Cumulative distribution function (black line) of ellipticity in grid cells recorded in a novel room. The optimal elongation factor is indicated in blue. Note the proximity of this factor to the median ellipticity level. **i**, Frequency histograms showing actual distribution of orientation for each grid axis (top), and distribution after corner-shearing using the transform in **f** (middle and bottom). Middle panel, shearing to one pair of anchoring corners; bottom, shearing to the other pair. After corner-shearing to the first pair, the offset was minimized to the same extent as in the simple shearing paradigm (middle, red asterisk). The offset could be abolished almost completely by corner shearing (peak symmetry offset after shearing 0° , kernel smoothed density curve with Gaussian kernel width 1.35° , 60% of the data distributed within $0-5^\circ$). After corner shearing, however, the distribution of ellipticity displayed less variation than after simple shearing (s.d. of 0.0025 versus 0.009).

j, Proposed model of deformation and rotation of grid patterns as a function of experience. From left to right, default minimum-energy state of grid, elongated grid in novel environment, sheared grid with rotation after experience, and reversal of sheared grid to default state by reverse corner-shearing analysis. We suggest that, in novel environments, grid cells may start out with an orientation that aligns one grid axis with a band along a wall defined by the activity of border cells. This initial alignment may disrupt the symmetry of the grid pattern. Through shearing, the grid may then be relaxed towards a lower-energy-state solution that is less dependent on the initial anchoring segments.



Extended Data Figure 8 | Distribution of offset and ellipticity in a circular environment. We have shown that, in square environments, the grid pattern is deformed and rotated by shear forces parallel to the walls of the environment. Here we show how grid orientation and grid ellipticity are distributed in a circular environment. We used a sample of 23 grid cells from 6 rats that had been recorded in a 2 m wide circular environment in a previous study². 100% of the cells had 6 grid fields or more. **a**, Example grid rate maps (top) and associated spatial autocorrelograms (bottom). **b**, Histogram of grid orientation for all 3 grid axes and across all cells recorded in the circle. Grid orientation was

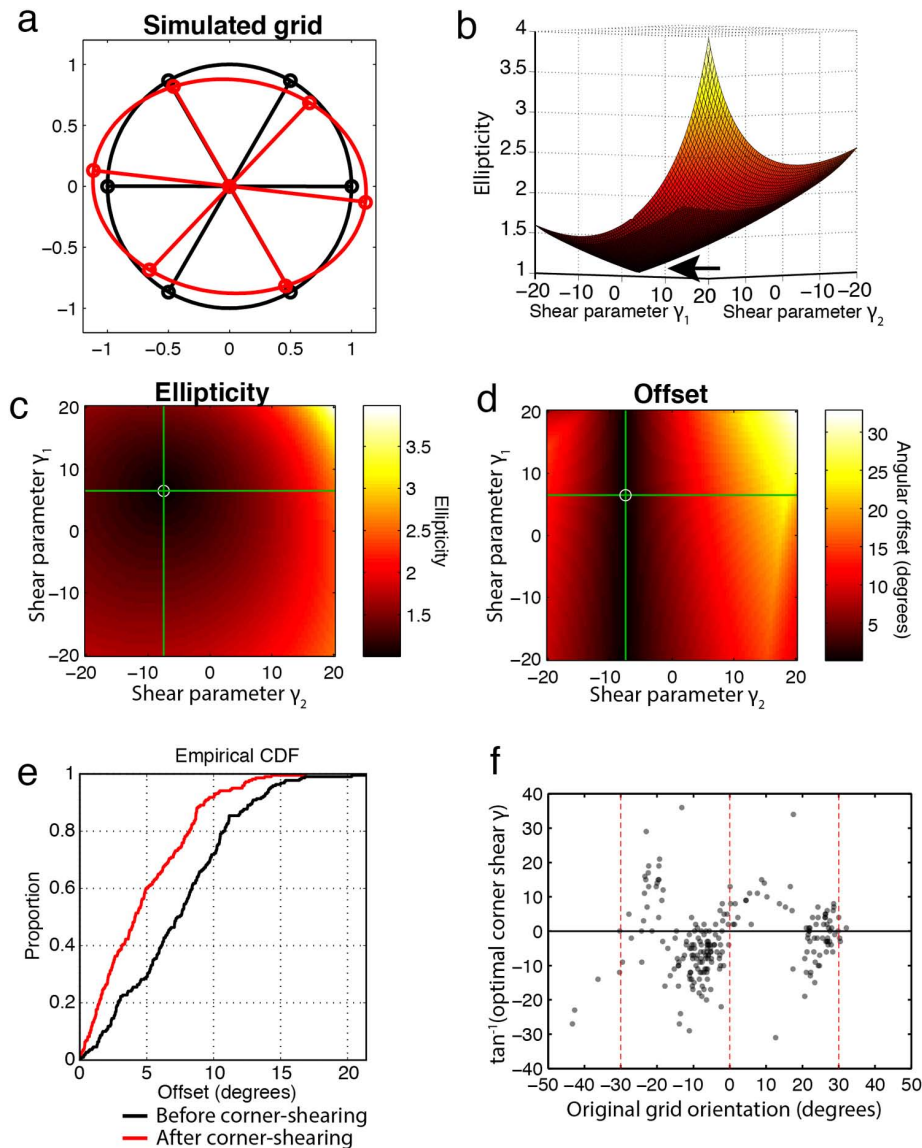
more variable than in the square environments. The distribution of mean grid orientation was not significantly different from uniform (Rayleigh test (mean grid orientation was multiplied by 6 to achieve 360° range); $Z = 1.23$, $P = 0.30$). The ellipticity of the grid pattern was significantly increased compared to the ellipticity in the square boxes (1.24 ± 0.025 versus 1.17 ± 0.004 ; $Z = -2.98$ $P = 0.003$). This increase in ellipticity likely reflects grid pattern variability due to the absence of local geometric landmarks such as corners.



Extended Data Figure 9 | Effect of shearing on deformation and reorientation of grid patterns in 1.5 m and 2.2 m environments.

a, b, Distribution of angular offset after shearing along each of the cardinal axes of the 2.2 m box. **a**, Kernel smoothed density estimates (Gaussian kernel width: 2.5°) of grid orientation for all 3 grid axes for all cells recorded in the 2.2 m box. The top panel shows orientation after shearing along the axis orthogonal to the alignment axis W_N (solid black curve). Shearing was performed separately for cells aligned to the east–west and north–south axis of the box and then combined on the basis of which relationship the shear axis had to the alignment axis (orthogonal for top plot). Dashed grey lines show the original distribution before shearing. Red lines indicate the orthogonal box axes ($0 \pm 30^\circ$, 0° corresponds to the east–west box axis, 30° to the north–south axis). Peaks detected in the sheared distribution around the red lines are reported above each peak. Note reduction—but not complete elimination—of angular offsets after shearing orthogonal to the alignment direction. Bottom, same as above but for shear axes parallel to the alignment axes. Shearing along this axis did not result in reduction of the offset. **b**, Kernel smoothed density curves (Gaussian kernel width: 1.35°) for absolute offsets from nearest multiples of 60° (referred to the parallel configuration) after shearing along both axes: orthogonal to alignment axis (top) and parallel to the alignment axis (bottom). Methods for shearing were the same as in **a**. Peak for each distribution is indicated by red solid lines. Axes were sorted per cell according to their angular distance from the nearest wall (axis 1 = A_{\min} ; axis 3 had the largest angle from nearest wall). After shearing orthogonal to the alignment axis, the offset peak was reduced considerably. Shearing along the other axes did not reduce the offset. **c–e**, The effect of adding a second shear interaction. **c**, Effects of anchoring to multiple walls in a simulated grid. A square lattice is used for illustrative purposes. Top row, square lattice aligned to the cardinal axes with no shear interactions from any wall (left). Simulated rate map of grid aligned to the east–west box axis (middle). Spatial autocorrelogram generated from rate map (right). Ellipse and inner fields are shown in black. Middle row, the square lattice after a shear transform along the north–south axis (left). Black line with arrow illustrates shearing interaction from wall. Shear origin (anchoring point) is shown in red. Simulated rate map after shearing is shown in the middle, and the resulting autocorrelogram is shown to the right. Bottom row: square lattice (left), simulated rate map (middle) and resulting autocorrelogram (right) after sequential shearing forces from two wall axes. Note that adding the second

shear force has no impact on the smallest angular offset (A_{\min}) but the orientation of axes 2 and 3 is changed and the ellipse orientation is altered accordingly. **d**, Grid axes and ellipses from autocorrelograms in **c** before (left) and after (right) the shearing that minimized both ellipticity and grid offset (illustrated with insets). The east–west axis and its 60° multiples are shown in grey. The axis nearest one of the walls is highlighted in red. Note the elimination of grid offset in the case with single shear interactions (middle), but inability of simple shearing to eliminate the offset (asterisk) when two shear interactions were operative (bottom). **e**, Effect of second shear interaction on grid orientation and deformation. We systematically applied shearing from a second shear axis to the simulated grid that was already sheared in one direction to induce a 7.5° offset. A range of shear factors was explored (-0.35 to 0.35). Left column, effect of second shear on grid orientation for individual grid axes. The middle panel shows the axis that was originally sheared to become offset by 7.5° . Note minimal change in offset in this axis, while systematic changes occurred in the two remaining axes (top and bottom respectively), resulting in further elliptic deformation of the grid pattern. Top right, grid orientation averaged across the 3 axes. Bottom right, ellipticity of the grid pattern as a function of the second shear parameter. **f**, Quadrant spatial autocorrelograms for grid cells with large spacing yielding fewer than 3 grid fields per quadrant. Left column, simulated rate maps with one, two or three grid fields. Middle column, spatial autocorrelograms from the rate maps in the left column. As expected, a single field was not informative in terms of grid features such as grid orientation or spacing, whereas two fields yielded information about one axis and three fields were informative about all axes. Taken together across all cells from a module, such contributions may average to overcome the effect of sparse field-sampling and recapitulate the original local grid geometry. Right column, simulation based on 20 grid cells with different grid phase (but similar spacing and orientation) showing that when spatial autocorrelograms from multiple maps (from one module) with varying number of fields from various axes are combined (averaged), grid features from all axes can be retrieved and used to determine grid geometry in subdivisions of the original environment. White asterisks show field peaks detected by the algorithm. Orange lines show the grid orientation used to generate the simulated grid cells within a quadrant. The average autocorrelogram faithfully captures the original grid geometry even if the average number of fields per cell is less than three.



Extended Data Figure 10 | Two-axis corner-shearing removed the offset in the 2.2 m environment. Several observations point to shear-like forces from two perpendicular walls as more common in the 2.2 m environment than the 1.5 m environment (Fig. 3). Minimizing ellipticity by simple shearing completely removed the angular offset in the 1.5 m box. In the 2.2 m box, however, the effect was only moderate, reflecting the fact that the grid was anchored to two walls, not one. Here we sought to determine if the two optimal shear parameters associated with two-axis corner-shearing were recoverable with the same approach as for simple shearing (minimization of ellipticity). Specifically, we tried to recover the original configuration of a simulated grid pattern on which we had applied two-axis corner-shearing in advance. The starting (default) grid pattern was made perfectly symmetrical and aligned to the east–west axis. We then applied 2-axis corner-shearing using the reverse transform proposed to occur as the result of grid shrinkage with corner anchoring during familiarization (corner-shearing):

$$f(x, y) = \begin{bmatrix} 1 + \gamma_1 & \gamma_1 \\ \gamma_2 & 1 + \gamma_2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

We set $\gamma_1 = -\frac{1}{2} \tan \frac{\pi}{24}$, and $\gamma_2 = \frac{1}{2} \tan \frac{13\pi}{360}$. **a**, Simulated grid pattern before (black) and after (red) two-axis shearing. The offset from the first shear axis is 7.5° but the second shear axis introduces further rotation, combined with deformation of the two remaining axes. **b**, Ellipticity surface resulting from systematic exploration of the two-shear parameter dimensions. Height and colour indicate ellipticity. x and y axes denote the angular offset that the shearing parameters would cause in simple shearing (from -40 to 40°). Note

the prominent minimum corresponding to a unique solution to the problem of recovering the original pair of shear parameters (arrow). **c**, Same as in **b**, but as pure colour map. The parameter set that was used to shear the grid initially is shown with green lines. The point of minimum ellipticity is shown as a white circle. In this case, the retrieved parameters were exactly those that were used in the original transform of the pattern. The example illustrates that the original shear parameter set could be recovered completely by two-axis reverse shearing. **d**, Colour map of angular offset resulting from two-axis shearing across the same range of shear parameter values as in **a**. Black corresponds to wall alignment (0° offset). The point of minimal ellipticity (white circle in **c**) is the same as for minimal offset (white circle in **d**). **e**, Cumulative distribution function showing angular offset of the data recorded in the 2.2 m box before (black) and after (red) two-axis corner shearing (Fig. 3h). Note consistent shift to the left after shearing. 2-axis corner shearing significantly reduced the original offset (Wilcoxon rank sum test: $Z = 6.6$, $P = 4.4 \times 10^{-11}$); the offset was abolished almost completely by corner shearing (peak symmetry offset after shearing 0° , kernel smoothed density curve with Gaussian kernel width 1.35° , 60% of the data distributed within 0 – 5°), while normal 2-axis shearing had little effect (Fig. 3h). **f**, Scatterplot of optimal corner-shear parameter and original grid offset for all data in the 2.2 m box. Shown is \tan^{-1} of the shear parameter (in degrees) to illustrate offset that the parameter would yield in simple shearing). The two main scatter clusters correspond to the two distinct anchoring solutions in the 2.2 m environment (that is, the two wall-axes). Red dashed lines show the wall axes of the environment ($[-30, 0, 30]^\circ$).

The formation of a quadruple star system with wide separation

Jaime E. Pineda¹, Stella S. R. Offner^{2,3}, Richard J. Parker⁴, Héctor G. Arce⁵, Alyssa A. Goodman⁶, Paola Caselli⁷, Gary A. Fuller⁸, Tyler L. Bourke^{9,10} & Stuart A. Corder^{11,12}

The initial multiplicity of stellar systems is highly uncertain. A number of mechanisms have been proposed to explain the origin of binary and multiple star systems, including core fragmentation, disk fragmentation and stellar capture^{1–3}. Observations show that protostellar and pre-main-sequence multiplicity is higher than the multiplicity found in field stars^{4–7}, which suggests that dynamical interactions occur early, splitting up multiple systems and modifying the initial stellar separations^{8,9}. Without direct, high-resolution observations of forming systems, however, it is difficult to determine the true initial multiplicity and the dominant binary formation mechanism. Here we report observations of a wide-separation (greater than 1,000 astronomical units) quadruple system composed of a young protostar and three gravitationally bound dense gas condensations. These condensations are the result of fragmentation of dense gas filaments, and each condensation is expected to form a star on a timescale of 40,000 years. We determine that the closest pair will form a bound binary, while the quadruple stellar system itself is bound but unstable on timescales of 500,000 years (comparable to the lifetime of the embedded protostellar phase¹⁰). These observations suggest that filament fragmentation on length scales of about 5,000 astronomical units offers a viable pathway to the formation of multiple systems.

Barnard 5 (B5) is a dense core in the Perseus star-forming region (at a distance of 250 pc) that hosts at least one young, forming star¹¹. Imaging of B5 in the emission of the dense-gas-tracing $\text{NH}_3(1,1)$ line shows it to be an example of a ‘coherent’ dense core¹², which is a contiguous high-density region with subsonic levels of turbulence¹³. Higher-resolution imaging reveals narrow filamentary structure within the coherent core¹⁴. We observed the $\text{NH}_3(1,1)$ and $(2,2)$ lines using the Karl G. Jansky Very Large Array (VLA)¹⁵, which reveals that the filaments in B5 are fragmenting and that they are in the process of forming a wide-separation multiple stellar system.

Nearly half of all stars reside in multiple star systems^{4,16}. Consequently, a host of phenomena, ranging from supernova rates to planet formation, depend on understanding stellar multiplicity¹⁷. Because of the observational challenges associated with observing early systems, the dominant ideas for binary formation are based on simulations and analytic arguments, which naturally require a variety of assumptions^{4,18}. To date, observations have not captured the formation of a binary system at a stage where its origin is unambiguous, and prior observations of core substructure lack the spatial and kinematic resolution to be used in predicting whether observed structures would form protostars and/or produce a bound system^{19,20}. The observed kinematics and separation ($>1,000$ astronomical units, AU) of the B5 system is significant because it demonstrates a clear mechanism for wide binary formation and provides convincing evidence that the observed condensations will become a bound multiple star system.

Detailed knowledge of the underlying distribution of dense gas is the key to determining which structures will go on to form stars. Here we identify the dense gas structures that are most likely to form stars using the dendrogram technique²¹. Dendrogram analysis is a hierarchical structure decomposition that uses isocontours to identify individual features, while also determining where these contours merge with adjacent structures to create a new parental structure. We refer to the smallest scale (and brightest) structures in the dendrogram as condensations. These are the most likely places for an individual star to form. Figure 1a shows the B5 region as seen in dense gas (number density of H_2 , $n_{\text{H}_2} \gtrsim 10^4 \text{ cm}^{-3}$), with the protostar and the identified gas condensations shown by a star and circles, respectively. The mass of the well known protostar B5-IRS1 is 0.1 solar masses (M_{Sun} ; ref. 22), while the masses of condensations B5-Cond1, B5-Cond2 and B5-Cond3 are $0.36 \pm 0.09 M_{\text{Sun}}$, $0.26 \pm 0.12 M_{\text{Sun}}$ and $0.30 \pm 0.13 M_{\text{Sun}}$, respectively. Uncertainty in these masses is dominated by the uncertainty in the temperature used to convert measured fluxes to masses. The radii of the three condensations are respectively 2,800 AU, 2,300 AU and 2,500 AU, while the projected separations between the same three condensations and the protostar are 3,300 AU, 5,100 AU and 11,400 AU (see Methods). The half-mass radii of the condensations are about half the condensation radii. This, combined with the mass radius relations (Extended Data Fig. 2), suggests that the central regions will collapse faster than the whole condensations and before interactions between condensations can play a major role in the stars’ initial separations. Although these separations are large, they are consistent with initial protostellar pair separations predicted for core fragmentation by numerical simulations². In the simulations, protostellar separations evolve rapidly on timescales of 0.1 Myr, and some systems become unbound while others migrate to closer proximity.

Projected proximity on the sky does not necessarily imply that objects are physically related. However, the line-of-sight velocities of the observed condensations are similar and the grouping is likely to be physically associated. The velocity dispersion, σ_v , of the dense gas provides another important piece of information, the gas kinetic energy, which is needed to determine whether the condensations are transient structures or gravitationally bound and likely to form a star. The velocities and velocity dispersions of the condensations are determined by fitting $\text{NH}_3(1,1)$ and $(2,2)$ line profiles¹⁴. The condensations and protostar display the same centroid velocity to within 0.2 km s^{-1} and are therefore associated with the same dense core. The level of turbulence in this region is so low, $\sigma_{\text{turb}} \approx 0.53\text{--}0.66$ times the sound speed in the gas¹⁴, that gravity will overwhelm the combined turbulent and thermal pressure in all the identified condensations, and a star will probably form in each case. The timescale for these condensations to undergo gravitational collapse is approximately the gas free-fall time, which we estimate to be 40,000 years (Methods). This timescale is sufficiently short to ensure that the system’s spatial

¹Institute for Astronomy, ETH Zurich, Wolfgang-Pauli-Strasse 27, CH-8093 Zurich, Switzerland. ²Department of Astronomy, Yale University, PO Box 208101, New Haven, Connecticut 06520-8101, USA. ³Department of Astronomy, University of Massachusetts, 710 North Pleasant Street, Amherst, Massachusetts 01003, USA. ⁴Astrophysics Research Institute, Liverpool John Moores University, 146 Brownlow Hill, Liverpool L3 5RF, UK. ⁵Department of Astronomy, Yale University, PO Box 208101, New Haven, Connecticut 06520-8101, USA. ⁶Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, Massachusetts 02138, USA. ⁷Max-Planck-Institut für extraterrestrische Physik (MPE), Gießenbachstrasse 1, D-85741 Garching, Germany. ⁸UK ARC Node, Jodrell Bank Centre for Astrophysics, School of Physics and Astronomy, Alan Turing Building, Oxford Road, University of Manchester, Manchester M13 9PL, UK. ⁹SKA Organisation, Jodrell Bank Observatory, Lower Withington, Macclesfield, Cheshire SK11 9DL, UK. ¹⁰Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, Massachusetts 02138, USA. ¹¹Joint ALMA Observatory, Alonso de Cordova 3107, Vitacura, Santiago, Chile. ¹²National Radio Astronomy Observatory, 520 Edgemont Road, Charlottesville, Virginia 22903, USA.

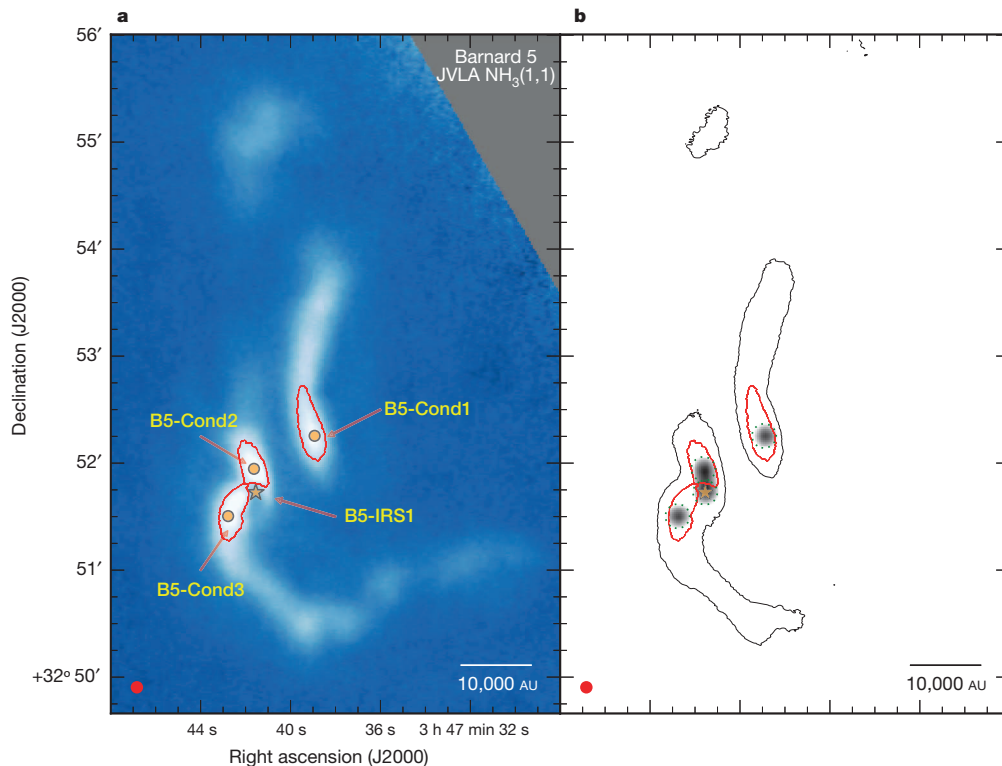


Figure 1 | High-angular-resolution image of dense gas and stellar progenitors. **a**, Background image from the JVLA of the Barnard 5 region shows the dense gas traced by $\text{NH}_3(1,1)$. It reveals two filaments, which together host three gravitationally bound condensations (B5-Cond1, 2, 3). Red contours and orange-filled circles show the condensation boundaries and centres, while the star

indicates the protostar (B5-IRS1) location. **b**, Contour map showing the filaments in dense gas. Greyscale circles show the distance that could be covered during 40,000 years while moving at the speed of sound, 0.2 km s^{-1} . In **a** and **b**, filled red circle at bottom left shows the angular resolution of the observations, and scale bar is shown at bottom right.

configuration will remain nearly unchanged during collapse, even if the protostars move as far as possible (in straight lines at the gas sound speed) for the duration of the collapse ($<1,700 \text{ AU}$ or $6.8''$ at the distance of Perseus). Figure 1b shows, as dotted circles, the possible ranges over which protostars could move in a free-fall time. The circles are smaller than the current condensation radii, so we conclude that this is a multiple system caught at the beginning of its formation.

In order to determine if the multiple stellar system is bound, it is necessary to estimate the masses of the stars that will be formed within the condensations. However, there are two complicating factors that make mass estimates uncertain. First, we must estimate what fraction of the condensation mass will end up in the star. Comparison between the initial mass function of stars and the distribution of dense core masses suggests that individual cores have a star formation efficiency of $\epsilon_{\text{DenseCore}} = M_{\text{Star}}/M_{\text{DenseCore}} \approx 30\%$ (refs 23–25). Theoretical estimates based on the effects of protostellar outflows predict core efficiencies of $\epsilon_{\text{DenseCore}} = 25\text{--}75\%$ (refs 26–28). Here, the condensations are embedded inside a previously identified dense core, B5, and they have radii one-tenth those of typical dense cores. Therefore, it would not be surprising if their star formation efficiency is close to 100%, at $\epsilon_{\text{Condensation}} = M_{\text{Star}}/M_{\text{Condensation}} \approx 75\%$ (ref. 26). Second, since the condensations are embedded within dense filaments, it is possible that the final stellar masses will be higher than the current measured condensation mass, because additional gas can flow along the filaments and accrete onto the condensations^{29,30}. If we adopt a very conservative estimate for the efficiency of 30%, and assume no additional mass accretion from the filaments, the final stellar masses formed from the condensations will be above the brown dwarf limit (80 Jupiter masses). Further fragmentation could occur within the condensations, however, given the low multiplicity fraction at these masses, $<26\%$ (ref. 4), so for our analysis we assume the most likely outcome, single stars. The lower the final stellar masses, the less likely the system is to be bound, so we

have made the most pessimistic assumptions possible (giving the lowest stellar masses) in evaluating boundedness.

Depending on the final stellar masses and kinematics, the resulting multiple stellar system could either be strongly bound or quickly dissolve owing to dynamical interactions. Given the dynamical instability of higher order systems, it is very likely that even if all the stars are initially bound, one or more will be dynamically ejected at a later time^{8,9}. For the closest possible pair, B5-IRS1 and B5-Cond2, we calculate the energies for a wide range of final stellar masses and under different kinematical and spatial distribution assumptions. In each case, we make different assumptions to reconstruct the total velocity dispersion and total separation based on our measured line-of-sight values. Figure 2 shows that the ratio of kinetic to gravitational energy is much less than one for all these cases, and therefore the pair is gravitationally bound.

Similarly, we compare the kinetic and gravitational energies for the expected stellar system taken all together, and find that the quadruple system is bound. Although bound, it is not likely to be a stable hierarchical system in the long term (see Methods), and the system will probably dissipate into a wide-separation binary system (B5-IRS1 and B5-Cond2). An important caveat is that this analysis does not take into account the effect of gas. The system is embedded in a larger reservoir of gas (the B5 core), which is several times the combined mass of the condensations. This additional gas can have two effects on the system evolution. First, the gravitational potential of the dense gas enhances the binding energy of the system by increasing the stellar escape velocity. Eventually, much of this gas will be removed by outflows. Second, the gas acts as a drag force on the stars, dissipating some of the stellar kinetic energy. Both these effects support the same outcome: a bound stellar system, at least during the formation stage. These results show that fragmentation of filaments would happen at scales smaller than those predicted by Jeans fragmentation of dense cores^{14,25}, and therefore filaments (or substructure in general) might be crucial ingredients in the formation of multiples.

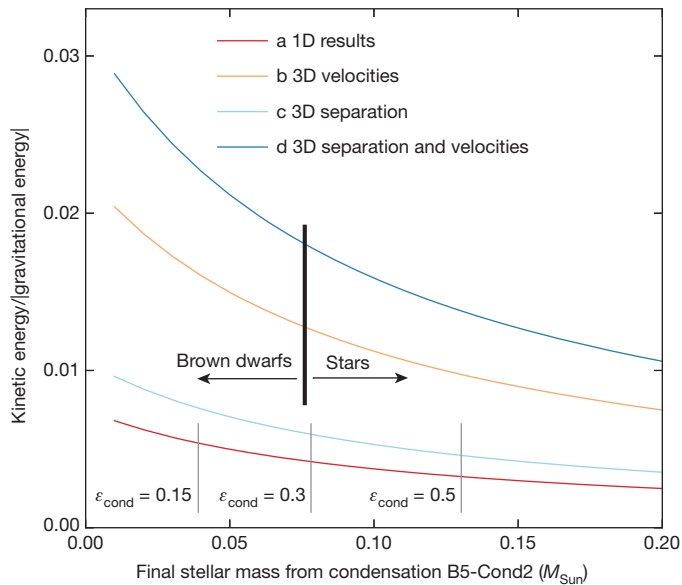


Figure 2 | Ratio of kinetic to gravitational energy for B5-IRS1 and B5-Cond2 as a function of final stellar mass. A binary system is bound if the kinetic-to-gravitational energy ratio is below unity. We calculate the energies for a wide range of efficiencies, $\epsilon_{\text{Condensation}} = M_{\text{Star}}/M_{\text{Condensation}}$, and under different kinematical and spatial distribution assumptions (see text for details) for the B5-IRS1 protostar and the final object from condensation B5-Cond2. First, we use the on-sky separation and the line-of-sight velocity difference as the total binary distance and velocity difference (line a, red). Next, we assume that the velocity difference along the line-of-sight is representative for the difference in the other directions, and therefore the total velocity difference is $\sqrt{3}$ times the velocity difference along the line-of-sight (line b, orange). Line c (light blue) assumes that the binary separation on the plane of the sky is a good estimate for the separation along the line-of-sight, and therefore the total separation is estimated as $\sqrt{2}$ times the separation on the plane of the sky. Finally, we compute the energies assuming both b and c together (line d, dark blue). Grey vertical lines denote some representative efficiencies, $\epsilon_{\text{Condensation}}$. Black vertical line marks the upper mass limit for a brown dwarf ($0.076 M_{\text{Sun}}$). This figure shows that for all estimates the closest separation binary is bound.

Additional observations of the distribution of dense gas in other regions using the VLA and/or Atacama Large Millimeter Array (ALMA) will determine the frequency of occurrence of filaments (or substructure) in dense cores and the distribution of separations of pre-stellar condensations at birth.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 18 September; accepted 24 December 2014.

1. Stamatellos, D. & Whitworth, A. P. The properties of brown dwarfs and low-mass hydrogen-burning stars formed by disc fragmentation. *Mon. Not. R. Astron. Soc.* **392**, 413–427 (2009).
2. Offner, S. S. R., Kratter, K. M., Matzner, C. D., Krumholz, M. R. & Klein, R. I. The formation of low-mass binary star systems via turbulent fragmentation. *Astrophys. J.* **725**, 1485–1494 (2010).
3. Moeckel, N. & Bate, M. R. On the evolution of a star cluster and its multiple stellar systems following gas dispersal. *Mon. Not. R. Astron. Soc.* **404**, 721–737 (2010).
4. Duchêne, G. & Kraus, A. Stellar multiplicity. *Annu. Rev. Astron. Astrophys.* **51**, 269–310 (2013).
5. Chen, X. *et al.* SMA observations of class 0 protostars: a high angular resolution survey of protostellar binary systems. *Astrophys. J.* **768**, 110 (2013).
6. Connelley, M. S., Reipurth, B. & Tokunaga, A. T. The evolution of the multiplicity of embedded protostars. I. Sample properties and binary detections. *Astron. J.* **135**, 2496–2525 (2008).

7. Connelley, M. S., Reipurth, B. & Tokunaga, A. T. The evolution of the multiplicity of embedded protostars. II. Binary separation distribution and analysis. *Astron. J.* **135**, 2526–2536 (2008).
8. Reipurth, B. & Mikkola, S. Formation of the widest binary stars from dynamical unfolding of triple systems. *Nature* **492**, 221–224 (2012).
9. Reipurth, B. *et al.* Multiplicity in early stellar evolution. Preprint at <http://arXiv.org/abs/1403.1907> (2014).
10. Evans, N. J., II *et al.* The Spitzer c2d legacy results: star-formation rates and efficiencies; evolution and lifetimes. *Astrophys. J.* **181**, 321–350 (2009).
11. Fuller, G. A. *et al.* Anatomy of the Barnard 5 core. *Astrophys. J.* **376**, 135–149 (1991).
12. Pineda, J. E. *et al.* Direct observation of a sharp transition to coherence in dense cores. *Astrophys. J.* **712**, L116–L121 (2010).
13. Goodman, A. A., Barranco, J. A., Wilner, D. J. & Heyer, M. H. Coherence in dense cores. II. The transition to coherence. *Astrophys. J.* **504**, 223–246 (1998).
14. Pineda, J. E. *et al.* Expanded Very Large Array observations of the Barnard 5 star-forming core: embedded filaments revealed. *Astrophys. J.* **739**, L2 (2011).
15. Perley, R. A., Chandler, C. J., Butler, B. J. & Wrobel, J. M. The expanded Very Large Array: a new telescope for new science. *Astrophys. J.* **739**, L1 (2011).
16. Lada, C. J. Stellar multiplicity and the initial mass function: most stars are single. *Astrophys. J.* **640**, L63–L66 (2006).
17. Parker, R. J. & Quanz, S. P. On the frequency of planetary systems around G dwarfs. *Mon. Not. R. Astron. Soc.* **436**, 650–658 (2013).
18. Tohline, J. E. The origin of binary stars. *Annu. Rev. Astron. Astrophys.* **40**, 349–385 (2002).
19. Kamazaki, T., Saito, M., Hirano, N. & Kawabe, R. Millimeter-wave interferometric study of the rho Ophiuchi A region. I. Small-scale structures of dust continuum sources. *Astrophys. J.* **548**, 278–287 (2001).
20. Nakamura, F., Takakuwa, S. & Kawabe, R. Substellar-mass condensations in prestellar cores. *Astrophys. J.* **758**, L25 (2012).
21. Rosolowsky, E. W., Pineda, J. E., Kauffmann, J. & Goodman, A. A. Structural analysis of molecular clouds: dendrograms. *Astrophys. J.* **679**, 1338–1351 (2008).
22. Brassfield, E. & Bourke, T. L. Submillimeter Array observations of embedded class I object Barnard 5 IRS1. *Bull. Am. Astron. Soc.* **43**, 340.09 (2011).
23. Motte, F., Andre, P. & Neri, R. The initial conditions of star formation in the rho Ophiuchi main cloud: wide-field millimeter continuum mapping. *Astron. Astrophys.* **336**, 150–172 (1998).
24. Alves, J., Lombardi, M. & Lada, C. J. The mass function of dense molecular cores and the origin of the IMF. *Astron. Astrophys.* **462**, L17–L21 (2007).
25. André, P. *et al.* From filamentary networks to dense cores in molecular clouds: toward a new paradigm for star formation. Preprint at <http://arXiv.org/abs/1312.6232> (2014).
26. Matzner, C. D. & McKee, C. F. Efficiencies of low-mass star and star cluster formation. *Astrophys. J.* **545**, 364–378 (2000).
27. Machida, M. N. & Hosokawa, T. Evolution of protostellar outflow around low-mass protostar. *Mon. Not. R. Astron. Soc.* **431**, 1719–1744 (2013).
28. Offner, S. S. R. & Arce, H. G. Investigations of protostellar outflow launching and gas entrainment: hydrodynamic simulations and molecular emission. *Astrophys. J.* **784**, 61 (2014).
29. Kirk, H. *et al.* Filamentary accretion flows in the embedded Serpens South protocluster. *Astrophys. J.* **766**, 115 (2013).
30. Peretto, N. *et al.* Global collapse of molecular clouds as a formation mechanism for the most massive stars. *Astron. Astrophys.* **555**, A112 (2013).

Acknowledgements We thank K. Todorov for discussions and comments that improved the paper. We acknowledge support from the Swiss National Science Foundation project CRSII2_141880 (J.E.P.), NSF CAREER award AST-0845619 (H.G.A.), NSF grant AST-0908159 (A.A.G.), ERC project PALs 320620 (P.C.) and NASA grant NNX09AB89G (T.L.B.). S.S.R.O. acknowledges support from NASA through Hubble Fellowship grant 51311.01 awarded by the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., for NASA, under contract NAS 5-26555. R.J.P. acknowledges support from the Royal Astronomical Society through a research fellowship. This research made use of *astrodendro* (<http://www.dendrograms.org>), *Astropy*, and *APLpy* (<http://aplpy.github.com>). The James Clerk Maxwell Telescope is operated by the Joint Astronomy Centre on behalf of the Science and Technology Facilities Council of the UK, the National Research Council of Canada, and (up to 31 March 2013) the Netherlands Organisation for Scientific Research. Additional funds for the construction of SCUBA-2 were provided by the Canada Foundation for Innovation. The National Radio Astronomy Observatory is a facility of the NSF operated under cooperative agreement by Associated Universities, Inc.

Author Contributions J.E.P. led the project and reduced the VLA data. J.E.P. and S.S.R.O. wrote the manuscript. R.J.P. conducted the stability analysis of the multiple system. All authors contributed to the VLA proposal, discussed the results and implications, and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.E.P. (jpineda@mpe.mpg.de).

METHODS

Observations and data reduction. *Very Large Array.* We conducted VLA observations of the B5 region on 2011 October 16–17 in D-array configuration, and on 2012 January 13–14 in CnD-array configuration (project 11B-101). We used the high-frequency K-band receiver and configured the WIDAR correlator to observe two basebands of 4 MHz bandwidth around the $\text{NH}_3(1,1)$ and $(2,2)$ lines. Each baseband is split into 8 adjacent spectral windows of 500 kHz bandwidth, with a channel separation of 0.049 km s^{-1} . The quasar 3C 48 was used as flux calibrator, 3C 84 as the bandpass calibrator, and J0336+3218 as the phase calibrator.

We reduced the data with the Common Astronomy Software Applications package³¹. The images were created using multi-scale clean (with scales of 0, 4 and 12 arcsec and *smallscalebias* parameter of 0.2) with a robust parameter of 0.5 and a 6 arcsec beam. We also included the NH_3 single dish data obtained with the Robert C. Byrd Green Bank Telescope¹² as a model image to recover the extended emission. The image has a noise level of 4 mJy per beam per channel.

The integrated emission map is calculated using all hyperfine components in the following velocity range: $(-10.016, -8.534) \text{ km s}^{-1}$, $(1.845, 3.575) \text{ km s}^{-1}$, $(9.012, 11.483) \text{ km s}^{-1}$, $(9.012, 11.483) \text{ km s}^{-1}$, $(16.919, 18.896) \text{ km s}^{-1}$, and $(28.781, 30.758) \text{ km s}^{-1}$. The final noise achieved is $2.8 \text{ mJy per beam km s}^{-1}$.

James Clerk Maxwell Telescope (JCMT). We observed B5 in the dust continuum emission at $450 \mu\text{m}$ and $850 \mu\text{m}$ using the Submillimetre Common-User Bolometer Array 2 (SCUBA-2) bolometer array³² at the JCMT. The observations were carried out on 2013 August 16 and 23, and on 2013 August 3 under project M13BU14 during grade 1 weather. We use the iterative map-making technique, *makemap*³³, with 0.5 arcsec pixels to match the $\text{NH}_3(1,1)$ VLA map. The initial reductions of the scans are co-added to create a mosaic, which we use to create a mask of signal-to-noise ratio >5 . Our final mosaic is created by co-adding a second reduction of the individual scans where the mask defines areas with emission. The angular resolution at $450 \mu\text{m}$ and $850 \mu\text{m}$ is 9.8 arcsec and 14.6 arcsec (ref. 34), respectively. We only use the $450 \mu\text{m}$ map, since its resolution is similar to the VLA observations.

We calibrate the flux scale of the observations using the Flux Correction Factor (FCF) of $4.71 \pm 0.5 \text{ Jy pW}^{-1} \text{ arcsec}^{-2}$ for $450 \mu\text{m}$ and of $2.34 \pm 0.08 \text{ Jy pW}^{-1} \text{ arcsec}^{-2}$ for $850 \mu\text{m}$ (ref. 34). We measure a noise level of $0.026 \text{ mJy per pixel}$ and $0.23 \text{ mJy per pixel}$ in the emission-free regions of the $850 \mu\text{m}$ and $450 \mu\text{m}$ maps. Extended Data Fig. 1 shows the final SCUBA-2 $450 \mu\text{m}$ and $850 \mu\text{m}$ maps, in which both the condensation B5-Cond1 and bright emission around the protostar B5-IRS1 are clearly detected, while condensations B5-Cond2 and B5-Cond3, although detected, are overwhelmed by the emission from B5-IRS1. The morphology of the emission is similar to the $\text{NH}_3(1,1)$ emission map, confirming the filamentary structure and the condensations.

Condensation identification. We use the dendrogram algorithm²¹ to identify the structures in the $\text{NH}_3(1,1)$ integrated intensity map. Each structure corresponds to a surface of constant intensity. An advantage of the algorithm is that structures can be nested inside one another, representing the hierarchy of the cloud, and then simply visualized using a tree diagram. The condensations are the ‘leaves’, that is, the highest level in the dendrogram decomposition. We produce the dendrogram with the following parameters: *min_value* = $4\sigma_{\text{rms}}$ (minimum intensity considered in the analysis), *min_delta* = $2\sigma_{\text{rms}}$ (minimum spacing between isocontours), and *min_npix* = 250 (minimum number of pixels contained in a structure), where $\sigma_{\text{rms}} = 5 \text{ mJy per beam km s}^{-1}$ is the rms noise. The condensation radius, R_{eq} , is defined as the equivalent radius, $\pi R_{\text{eq}}^2 = \text{area}$.

The centroid velocity of each condensation is computed by averaging the centroid velocity within a $6 \text{ arcsec} \times 6 \text{ arcsec}$ box centred at the peak emission in the $\text{NH}_3(1,1)$ map. The velocity dispersion of each condensation is determined by taking the average of the velocity dispersions within the condensation. In both cases, the uncertainties are estimated from the standard deviation of the quantities measured.

The total flux for each condensation is defined as the total flux minus the background emission removed within the structure boundary determined by the dendrogram. Extended Data Table 1 displays the fluxes calculated from the $\text{NH}_3(1,1)$ emission map for all condensations. Of the three condensations, only B5-Cond1 flux is not contaminated by the bright B5-IRS1 in the SCUBA-2 $450 \mu\text{m}$ dust emission continuum, and we include its total flux in the table.

Mass determination for condensations. We determine the total mass of B5-Cond1 using the total flux of the SCUBA-2 $450 \mu\text{m}$ dust emission map assuming optically thin emission,

$$M_{\text{dust}} = d^2 F_{\nu} / (\kappa_{\nu} B_{\nu}(T))$$

where d is the distance, F_{ν} is the total flux, κ_{ν} is the dust opacity per unit mass at frequency ν , and $B_{\nu}(T)$ is the Planck function at temperature T . We use a dust to gas ratio of 0.01, and a dust opacity per unit mass of $\kappa_{\nu} = 0.1 (\nu/1 \text{ THz})^2$ (ref. 35), which at $450 \mu\text{m}$ gives $\kappa_{450\mu\text{m}} = 0.044 \text{ cm}^2 \text{ g}^{-1}$. We assume a temperature of $T = 10 \text{ K}$,

which is the typical temperature of starless cores in Perseus³⁶, and distance of 250 pc (ref. 37). In the case of B5-Cond1, this gives a mass estimate of $0.39 M_{\text{Sun}}$.

We use the mass-to- NH_3 flux ratio derived from B5-Cond1, whose isolation makes mass calculations most reliable, to estimate the masses of condensations B5-Cond2 and B5-Cond3. This gives masses of $0.33 M_{\text{Sun}}$ and $0.39 M_{\text{Sun}}$ for B5-Cond2 and B5-Cond3, respectively.

We also plot the enclosed mass as a function of radius in Extended Data Fig. 2a. Here we also compare these mass–radius relations with those predicted for density profile of $\rho \propto r^{-2}$ and $\rho \propto r^{-1.5}$, and it is clear that the condensations are better described by a $\rho \propto r^{-1.5}$ density profile between 800 AU and $1,900 \text{ AU}$. A density profile of $\rho \propto r^{-2}$ is a good description for a core in (or close to) hydrostatic equilibrium, while a profile of $\rho \propto r^{-1.5}$ is a good model for free falling envelopes³⁸. This also supports the notion that the condensations in B5 are centrally concentrated and close to forming a central object.

Virial analysis for condensations. We calculate the virial parameter (the ratio between kinetic and gravitational energy) for a spherical core assuming a density profile $\rho \propto r^{-1.5}$ (ref. 39), which (see above) is a good approximation for the condensation’s density profile from 800 AU to $1,900 \text{ AU}$ (see Extended Data Fig. 2):

$$\alpha = 4 R \sigma_v^2 / G M = 928 (R/\text{pc}) (M/M_{\text{Sun}})^{-1} (\sigma_v/\text{km s}^{-1})^2$$

where R is the radius, M is the mass, and σ_v is the total velocity dispersion of the gas (including thermal and non-thermal components, $\sigma_v^2 = \sigma_{\text{th}}^2 + \sigma_{\text{turb}}^2$). A condensation is bound if $\alpha < 2$.

The average level of turbulence, σ_{turb} , measured towards condensations B5-Cond1, B5-Cond2 and B5-Cond3 is $0.66 \times \sigma_{\text{th}}$, $0.58 \times \sigma_{\text{th}}$ and $0.54 \times \sigma_{\text{th}}$, respectively, where $\sigma_{\text{th}} = 0.2 \text{ km s}^{-1}$ is the sound speed of the gas with mean molecular weight per free particle for $\mu = 2.33$ and at a temperature of 10 K .

We calculate the virial parameter as a function of radius for each condensation using its measured average velocity dispersion (Extended Data Fig. 2b); the condensations are all bound beyond a radius of $1,200 \text{ AU}$. The virial parameter initially decreases as a function of radius until it reaches a minimum at $\sim 60\%$ of the condensation radius, and then it increases until it reaches the edge of the condensation as defined by the dendrogram, while staying bound (Extended Data Fig. 2b).

Free-fall timescale. The free-fall timescale for a uniform density sphere is defined as

$$t_{\text{ff}} = \sqrt{\frac{3\pi}{32G\rho}} = \sqrt{\frac{\pi^2 R^3}{8GM}} = 1.6 \times 10^6 \left(\frac{R}{0.1 \text{ pc}} \right)^{3/2} \left(\frac{M}{0.1 M_{\text{Sun}}} \right)^{-1/2} \text{ yr}$$

where $M = 4\pi R^3 \rho/3$. The free-fall timescale, t_{ff} , is $4.0 \times 10^4 \text{ yr}$, $3.5 \times 10^4 \text{ yr}$, and $4.1 \times 10^4 \text{ yr}$ for B5-Cond1, B5-Cond2 and B5-Cond3, respectively. Thus, about 40,000 years is the ‘typical’ free-fall timescale for this system. This definition excludes the influence of magnetic fields, which could delay collapse.

Stability analysis of the multiple system. We calculate the potential and kinetic energy of each object and determine if all of the objects are bound⁴⁰. The gravitational potential energy, V_{p} , is calculated as

$$V_{\text{p}} = - \sum_{i \neq j} \frac{G m_i m_j}{r_{ij}}$$

where m_i and m_j are the masses of objects i and j , and r_{ij} is the distance between them. The kinetic energy of each object, T_{p} , is given by

$$T_{\text{p}} = \frac{1}{2} m_i (v_i - v_{\text{com}})^2$$

where v_i is the (line-of-sight) velocity of object i , and v_{com} is the velocity of the centre of mass of the system. A star is bound to the system if $T_{\text{p}}/|V_{\text{p}}| < 1$.

Using the stellar mass estimates, two-dimensional positions, and one-dimensional line-of-sight velocities, we find that all four objects comprise a bound system. The calculated kinetic-to-gravitational energy ratio of the system is 0.11. In order to assess the robustness of this result, we remove each object in turn from the analysis and find that the system remains bound. If we were to increase the mass of any of the four objects, this would decrease the energy ratio further.

We next determine whether any of the four objects will become a stable binary or triple system, with the caveat that we only know the velocity along the line-of-sight and the separation on the sky. We determine the binding energy, semimajor axis, and eccentricity of the central or ‘inner’ system (B5-IRS1 + B5-Cond2). We determine the eccentricity using the following formula⁴¹:

$$e = \sqrt{1 + \frac{2(m_1 + m_2)L_{\text{tot}}^2 E_{\text{bind}}}{G^2 m_1^3 m_2^3}}$$

where m_1 and m_2 are the component masses, L_{tot} is the magnitude of the total angular momentum vector on the centre of mass frame, and E_{bind} is the binding energy.

The angular momentum is calculated using only the line-of-sight velocity and the projected on-sky separation between the considered components, $L_{\text{tot}} = [m_1 m_2 / (m_1 + m_2)] [(x_2 - x_1, y_2 - y_1, 0) \times (0, 0, v_2 - v_1)]$, where x_i and y_i are the on-sky relative positions and v_i are the line-of-sight velocities. The binding energy is negative, and the semimajor axis is $a_{\text{in}} = 1,666$ AU with a high eccentricity of 0.99. We note that changing the component masses has a minimal effect on the semimajor axis and eccentricity values (if all the masses are $0.2 M_{\text{Sun}}$, then $a_{\text{in}} = 1,663$ AU, $e_{\text{in}} = 0.99$ and the system remains bound). However, the determination of a high eccentricity must be taken with caution, because its uncertainty is mostly due to the lack of knowledge about the full orbital elements.

Next, we treat the central system as the inner orbit of a triple system and determine whether either [(B5-IRS1 + B5-Cond2) + B5-Cond3] or [(B5-IRS1 + B5-Cond2) + B5-Cond1] could be a stable triple system. The former system could be a stable triple because it has a negative binding energy; the outer semimajor axis of this system is $a_{\text{out}} = 4,060$ AU, the eccentricity is 0.50, and the inner and outer orbit periods are 2×10^5 yr and 5×10^5 yr, respectively.

However, we note that as the on-sky separation of the inner system (B5-IRS1 + B5-Cond2) is less than 5–10 times the separation of the outer system [(B5-IRS1 + B5-Cond2) + B5-Cond3], which makes it highly likely to be unstable over long timescales⁸.

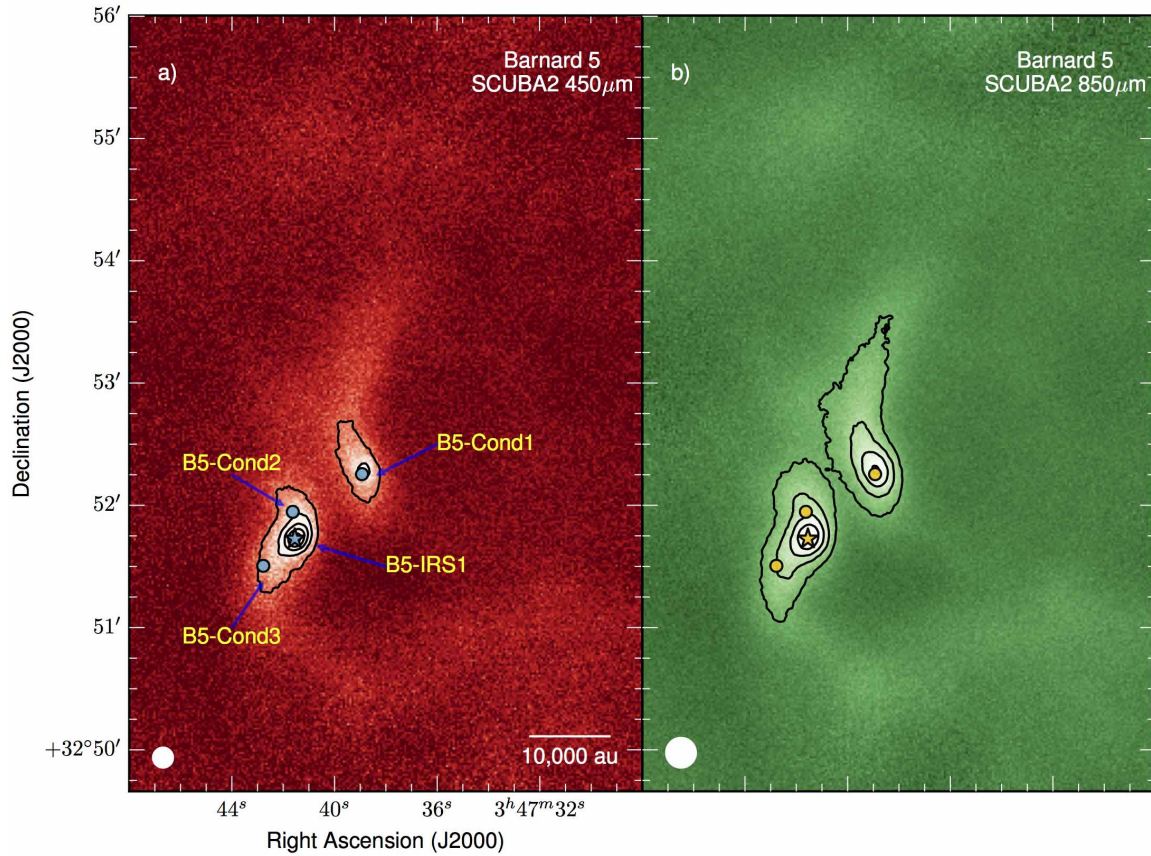
Finally, we find that the [(B5-IRS1 + B5-Cond2) + B5-Cond1] system has a positive binding energy, and is therefore not a stable bound system.

In summary, the four objects together are bound, but they do not constitute a stable hierarchical quadruple system. The central (B5-IRS1 + B5-Cond2) system is a bound binary (albeit with high eccentricity) and may make up a triple system with [(B5-IRS1 + B5-Cond2) + B5-Cond3].

The fate of this system will probably be determined by internal evolution, that is, whether or not it becomes unstable, rather than by external perturbations⁴¹, owing to it forming in a low stellar density environment.

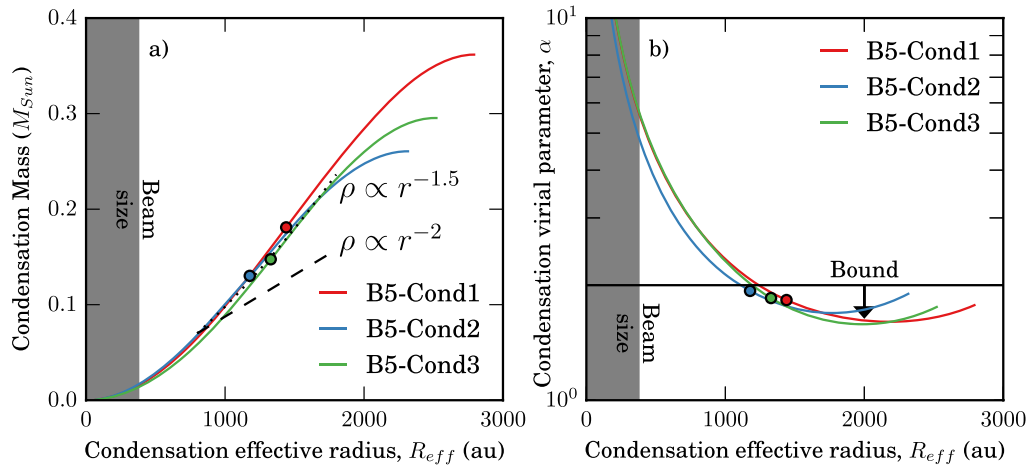
Code availability. The code used in this research is freely available at https://github.com/jpinedaf/B5_wide_multiple, it makes use of Astropy⁴².

31. McMullin, J. P., Waters, B., Schiebel, D., Young, W. & Golap, K. in *Astronomical Data Analysis Software and Systems XVI* (eds Shaw, R. A., Hill, F. & Bell, D. J.) 127–130 (Astron. Soc. Pacif. Conf. Ser. Vol. 376, 2007).
32. Holland, W. S. *et al.* SCUBA-2: the 10,000 pixel bolometer camera on the James Clerk Maxwell Telescope. *Mon. Not. R. Astron. Soc.* **430**, 2513–2533 (2013).
33. Chapin, E. L. *et al.* SCUBA-2: iterative map-making with the Sub-Millimetre User Reduction Facility. *Mon. Not. R. Astron. Soc.* **430**, 2545–2573 (2013).
34. Dempsey, J. T. *et al.* SCUBA-2: on-sky calibration using submillimetre standard sources. *Mon. Not. R. Astron. Soc.* **430**, 2534–2544 (2013).
35. Hildebrand, R. H. The determination of cloud masses and dust characteristics from submillimetre thermal emission. *Q. J. R. Astron. Soc.* **24**, 267–282 (1983).
36. Foster, J. B. *et al.* Dense cores in Perseus: the influence of stellar content and cluster environment. *Astrophys. J.* **696**, 298–319 (2009).
37. Enoch, M. L. *et al.* Bolocam survey for 1.1 mm dust continuum emission in the c2d legacy clouds. I. Perseus. *Astrophys. J.* **638**, 293–313 (2006).
38. Shu, F. H. Self-similar collapse of isothermal spheres and star formation. *Astrophys. J.* **214**, 488–497 (1977).
39. Enoch, M. L. *et al.* The mass distribution and lifetime of prestellar cores in Perseus, Serpens, and Ophiuchus. *Astrophys. J.* **684**, 1240–1259 (2008).
40. Baumgardt, H., Hut, P. & Heggie, D. C. Long-term evolution of isolated N-body systems. *Annu. Rev. Astron. Astrophys.* **336**, 1069–1081 (2002).
41. Parker, R. J. & Meyer, M. R. Binaries in the field: fossils of the star formation process? *Mon. Not. R. Astron. Soc.* **442**, 3722–3736 (2014).
42. Astropy Collaboration *et al.* Astropy: A community Python package for astronomy. *Astron. Astrophys.* **558**, A33 (2013).



Extended Data Figure 1 | Dust continuum emission maps of Barnard 5. **a, b,** Dust continuum emission observed with SCUBA-2 at 450 μm (**a**) and 850 μm (**b**). Contour levels are drawn at 3, 6, 9, 12 and 15 \times rms, where rms is 0.23 mJy per pixel and 0.026 mJy per pixel in the 450 μm and 850 μm maps, respectively. Emission from dust associated with the protostar B5-IRS1 dominates the field, and it makes it difficult to extract the emission from

B5-Cond2 and B5-Cond3. B5-Cond1 is clearly detected in the dust continuum emission. Since the dust continuum emission does show the presence of the filaments in low level emission, we conclude they are real column density features. Filled white circles at bottom left corners show the angular resolution of the observations. Blue- and orange-filled circles show the condensation centres, while the filled stars indicate the protostar (B5-IRS1) location.



Extended Data Figure 2 | Mass and virial parameter as a function of radius for condensations. **a**, The enclosed condensation mass, derived from $\text{NH}_3(1,1)$, at different effective radii for each condensation; **b**, the corresponding virial parameter as a function of effective radius for each condensation. The condensation mass grows rapidly with radius, with a profile similar to one expected for a density distribution of $\rho \propto r^{-1.5}$ (dotted line in **a**) until it is close to the condensation boundary. In comparison, the dashed line shows the expected result in hydrostatic equilibrium ($\rho \propto r^{-2}$), which is

different to the observed distribution. The virial parameter decreases with radius until it reaches a minimum of ~ 1.5 , and then it slowly increases until it reaches the condensation boundary. Notice that virial parameters below the horizontal line ($\alpha = 2$) imply bound condensations. The grey shaded region marks the regime where the effective radius is smaller than the angular resolution of the observations. The circles show the values at the half-mass radius.

Extended Data Table 1 | Condensation and protostar parameters

Source	RA	Dec	V_{lsr}	F_{NH_3}	F_{450}
	(hh:mm:ss.sss)	(dd:mm:ss.ss)	(km s^{-1})	(Jy km s^{-1})	(Jy)
B5-IRS1	03:47:41.548	32:51:43.57	10.21 ± 0.04	—	—
B5-Cond1	03:47:38.928	32:52:15.31	10.43 ± 0.03	0.349 ± 0.001	0.992 ± 0.009
B5-Cond2	03:47:41.627	32:51:56.81	10.23 ± 0.01	0.251 ± 0.001	—
B5-Cond3	03:47:42.778	32:51:30.31	10.30 ± 0.01	0.285 ± 0.001	—

RA, right ascension; Dec., declination; V_{lsr} , central velocity of the observed molecular line; F_{NH_3} , flux measured in the $\text{NH}_3(1,1)$ integrated intensity map; F_{450} , flux measured in the SCUBA-2 450 μm map.

Comet 67P/Churyumov–Gerasimenko sheds dust coat accumulated over the past four years

Rita Schulz¹, Martin Hilchenbach², Yves Langevin³, Jochen Kissel², Johan Silen⁴, Christelle Briois⁵, Cecile Engrand⁶, Klaus Hornung⁷, Donia Baklouti³, Anaïs Bardyn^{5,8}, Hervé Cottin⁸, Henning Fischer², Nicolas Fray⁸, Marie Godard⁶, Harry Lehto⁹, Léna Le Roy¹⁰, Sihane Merouane², François-Régis Orthous-Daunay¹¹, John Paquette², Jouni Rynö⁴, Sandra Siljeström¹², Oliver Stenzel², Laurent Thirkell⁵, Kurt Varmuza¹³ & Boris Zaprudin⁹

Comets are composed of dust and frozen gases. The ices are mixed with the refractory material either as an icy conglomerate¹, or as an aggregate of pre-solar grains (grains that existed prior to the formation of the Solar System), mantled by an ice layer^{2,3}. The presence of water-ice grains in periodic comets is now well established^{4–6}. Modelling of infrared spectra obtained about ten kilometres from the nucleus of comet Hartley 2 suggests that larger dust particles are being physically decoupled from fine-grained water-ice particles that may be aggregates⁷, which supports the icy-conglomerate model. It is known that comets build up crusts of dust that are subsequently shed as they approach perihelion^{8–10}. Micrometre-sized interplanetary dust particles collected in the Earth's stratosphere and certain micrometeorites are assumed to be of cometary origin^{11–13}. Here we report that grains collected from the Jupiter-family comet 67P/Churyumov–Gerasimenko come from a dusty crust that quenches the material outflow activity at the comet surface¹⁴. The larger grains (exceeding 50 micrometres across) are fluffy (with porosity over 50 per cent), and many shattered when collected on the target plate, suggesting that they are agglomerates of entities in the size range of interplanetary dust particles. Their surfaces are generally rich in sodium, which explains the high sodium abundance in cometary meteoroids¹⁵. The particles collected to date therefore probably represent parent material of interplanetary dust particles. This argues against comet dust being composed of a silicate core mantled by organic refractory material and then by a mixture of water-dominated ices^{2,3}. At its previous recurrence (orbital period 6.5 years), the comet's dust production doubled when it was between 2.7 and 2.5 astronomical units from the Sun¹⁴, indicating that this was when the nucleus shed its mantle. Once the mantle is shed, unprocessed material starts to supply the developing coma, radically changing its dust component, which then also contains icy grains, as detected during encounters with other comets closer to the Sun^{4,5}.

Since August 2014, the ESA Comet Rendezvous Mission, Rosetta^{16,17}, has been in orbit around the Jupiter-family comet 67P/Churyumov–Gerasimenko, monitoring the evolution of the comet's nucleus, near-nucleus region, and inner coma as a function of increasing solar flux input, as the comet moves towards the Sun. As part of these studies, the COmetary Secondary Ion Mass Analyser (COSIMA)¹⁸ onboard Rosetta is collecting comet grains from the near-nucleus region and the inner coma onto special target plates¹⁹, which are subsequently imaged and compositionally investigated by time-of-flight secondary ion mass spectrometry using an indium ion source. The grain collection commenced

at a heliocentric distance of 3.57 astronomical units (where 1 AU is the average Sun–Earth distance), when the comet was still at low activity. The optical analysis of the grains captured on the target plates at distances beyond 3 AU shows that most have fragmented upon capture and a large fraction of grains more than 50 μm across have shattered. Figure 1a shows a typical example of a dust particle that has crumbled into a rubble pile upon collection, while Fig. 1b shows an example of a dust particle that has shattered into a loosely connected cluster with a wide range of sub-component sizes. These two types of feature are representative of most large particles collected at less than 30 km from the nucleus during the first three months of the orbital phase. Given that the dust particles hit the target with a relatively low velocity ($1\text{--}10\text{ m s}^{-1}$)¹⁹, their tensile strength must be very low. From the inertial deceleration forces upon grain capture the strength of the material can be approximated, and a first rough estimate relevant for the present fragmentation process is on the order of 1,000 Pa.

The disintegration of cometary grains in the coma is often described as resulting from an icy grain component that evaporates when exposed to solar radiation, producing a secondary source for comet gaseous material^{20,21}. A dusty secondary source can, however, also be attributed to certain organic grains that are not mantled by water ice²². The coma dust returned by Stardust²³ featured various types of grain, including specimens that had disintegrated along the deceleration tracks when entering the aerogel (the ultralight porous gel in which the grains were captured) at velocities of the order of 6 km s^{-1} , and hence were composed of very fine or thermally unstable components^{24,25}. The morphology of the grains collected by COSIMA supports the presence of solely refractory material. A grain composed of an ice–mineral mixture would not shatter at low-velocity collection; instead, the icy part of such a grain would evaporate very shortly after collection, leaving one or more voids in the particle that remains on the target plate. Grains composed of (nearly) pure water-ice would evaporate at or shortly after collection and create a dark signature on the target plate. At the scale of the COSIMA image resolution (pixel size is $14\text{ }\mu\text{m}$), there is no hint of volatiles having left the grains after collection. In other words, there is no indication of an ice–mineral mixture, or of pure icy grains hitting the target. This is in contrast to cometary grains remotely observed, or collected before the Rosetta mission.

The most important difference between the Stardust and COSIMA grains is the heliocentric distance at which they were captured. The Stardust samples were collected during a comet fly-by at 1.85 AU, whereas the grains collected by COSIMA were dragged off the nucleus of a

¹European Space Agency, Scientific Support Office, Keplerlaan 1, Postbus 299, 2200 AG Noordwijk, The Netherlands. ²Max-Planck-Institut für Sonnensystemforschung, Justus-von-Liebig-Weg 3, 37077 Göttingen, Germany. ³Institut d'Astrophysique Spatiale, CNRS/Université Paris Sud, Bâtiment 121, 91405 Orsay, France. ⁴Finnish Meteorological Institute, Observation services, Erik Palménin aukio 1, FI-00560 Helsinki, Finland. ⁵Laboratoire de Physique et Chimie de l'Environnement et de l'Espace (LPC2E), CNRS/Université d'Orléans, 45071 Orléans, France. ⁶Centre de Sciences Nucléaires et de Sciences de la Matière, CNRS/IN2P3—Université Paris Sud—UMR8609, Bâtiment 104, 91405 Orsay campus, France. ⁷Universität der Bundeswehr, LRT-7, Werner Heisenberg Weg 39, 85577 Neubiberg, Germany. ⁸Laboratoire Interuniversitaire des Systèmes Atmosphériques (LISA), UMR CNRS 7583, Université Paris Est Créteil et Université Paris Diderot, Institut Pierre Simon Laplace, 94000 Créteil, France. ⁹University of Turku, Department of Physics and Astronomy, Tuorla Observatory Väisäläntie 20, 21500 Piikkiö, Finland. ¹⁰Center for Space and Habitability (CSH), University of Bern, Sidlerstrasse 5, 3012 Bern, Switzerland. ¹¹Université Grenoble Alpes/CNRS, Institut de Planétologie et d'Astrophysique de Grenoble, 414 Rue de la Piscine, Domaine Universitaire, 38000 Grenoble, France. ¹²Department of Chemistry, Materials and Surfaces, SP Technical Research Institute of Sweden, Box 857, 50115 Borås, Sweden. ¹³Institut für Statistik und Wahrscheinlichkeitstheorie, Technische Universität Wien, Wiedner Hauptstrasse 7, 1040 Wien, Austria.

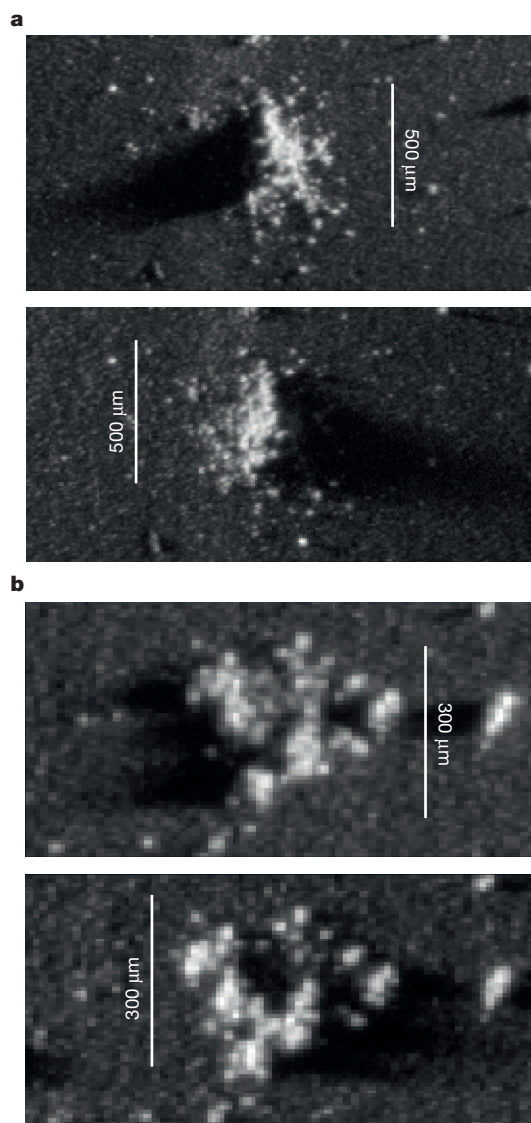


Figure 1 | Dust particles. **a**, An example of a dust particle that crumbled into a rubble pile when collected. The particle was collected at a nucleus distance of 10–20 km, between 25 and 31 October 2014, with corresponding heliocentric distance range 3.11–3.07 AU. The image was obtained with two different grazing illumination conditions (top image illuminated from the right, bottom image from the left). The brightness is presented in logarithmic scale to emphasize the shadows, which indicate that the altitude above the target reaches about 100 μm . As the particle lies 4.2 mm below the centre of the collecting target, the shadows are tilted with regard to the horizontal direction. **b**, An example of a dust particle that shattered when collected. The distance, time of collection, illumination conditions, and logarithmic scale are the same as for **a**. The shadows indicate that the altitude above the target reaches about 60 μm . The two grains visible on the right are not part of the shattered cluster.

re-approaching comet at heliocentric distances greater than 3 AU (as 67P/Churyumov-Gerasimenko returned from its aphelion passage at 5.68 AU having spent about four years at a distance beyond 4 AU). These COSIMA grains therefore come from a dusty layer that has built up over those four years, when the comet was so far from the Sun that the solar radiation was no longer able to create a gas drag that could efficiently remove the dust. The dust therefore remained on the surface, building up an ice-free, fluffy layer, below which lies an ice–dust mixture. When the comet returned to regions of higher solar irradiation the evaporation rate of the volatile gases underneath the dust layer increased again, lifting the particles from the dry upper dust layer into the inner coma,

and leaving their original dusty cohabitants (dust frozen together with the gas) behind. This left-behind dust replenishes the existing dusty layer from below, thereby maintaining its thickness in a quasi-steady state until the solar radiation is high enough that the amount of dust removed from the upper layer is larger than the new volatile-free dust produced underneath. As a consequence, the dusty layer will disappear over time and fresh material will come to the surface. The transition may be gradual but could be violent if there is a hard zone under the dusty layer (as may be indicated by the re-bounce of the Philae lander) below which high gas pressures are building up. From the increase in dust production rate observed telescopically in 2008 (ref. 14) we infer that the dusty layer was lost at some stage between 2.7 AU and 2.5 AU. That orbital section will be reached again during the present recurrence of the comet between 24 December 2014 and 20 January 2015, so the loss of the dusty layer has probably already occurred.

The mass spectra of the surface of the COSIMA grains collected beyond 3 AU show a high abundance of sodium. Preliminary values obtained after calibration²⁶ are as high as 0.8, normalized to Mg = 1. For comparison, the Na abundances (Mg = 1) for comet 81P/Wild-2 are 0.13 (collected in aerogel) and 0.2 (collected on aluminium foil)²⁷, 0.1 ± 0.06 for comet 1P/Halley²⁸, and 0.055 for CI chondrites²⁹. The Na abundance observed in Perseid and Leonid meteoroids is a factor of 1.5 higher than the chondritic value¹⁵, which fits very well with the value measured by COSIMA. Furthermore, the fluffiness of the COSIMA grains suggests that they would fragment with time after release into the coma. From remote observations, such fragmentation of coma grains has regularly been proposed³⁰. Therefore we conclude that the high Na abundance measured by COSIMA, combined with the fluffiness of the grains, supports the hypothesis that these grains represent the parent population of interplanetary dust particles in meteor streams of cometary origin.

Beyond 3 AU, COSIMA has not collected any of the dust that is mixed with sublimating ice, but rather the dust that is present in the upper ice-free dust layer. When the comet loses its fluffy mantle, it is expected that the properties of the grains collected will be very different from those of the grains currently under analysis, which show the properties of ‘space-weathered’ comet refractory material. The fresh material is likely to be a mixture of ice and dust, and its analysis should provide the detailed structure of this mixture. However, when the comet returns to the outer Solar System, a new dusty mantle will form as the upper layer again becomes free of ice. The formation of such a mantle was considered for re-occurring comets⁸ and detailed models exist for short-period comet nuclei^{9,10}. The physical processes and timescales of these models are consistent with assumptions made about the nucleus size, orbit and so on for 67P/Churyumov-Gerasimenko. Therefore, the grains collected from this comet provide direct evidence for the existence of its dusty mantle and also an indication of the structure of dust mantles in short-period comets.

Received 3 December; accepted 22 December 2014.

Published online 26 January; corrected online 11 February 2015 (see full-text HTML version for details).

- Whipple, F. L. A comet model. I. The acceleration of comet Encke. *Astrophys. J.* **111**, 375–394 (1950).
- Greenberg, J. M. Making a comet nucleus. *Astron. Astrophys.* **330**, 375–380 (1998).
- Greenberg, J. M. & Li, A. Morphological structure and chemical composition of cometary nuclei and dust. *Space Sci. Rev.* **90**, 149–161 (1999).
- Schulz, R. *et al.* Detection of water ice grains after the Deep Impact onto comet 9P/Tempel 1. *Astron. Astrophys.* **448**, L53–L56 (2006).
- A'Hearn, M. F. *et al.* EPOXI at Comet Hartley 2. *Science* **332**, 1396–1400 (2011).
- Yang, B., Jewitt, D. & Bus, S. J. Comet 17P/Holmes in outburst: the near infrared spectrum. *Astron. J.* **137**, 4538–4546 (2009).
- Protopapa, S. *et al.* Water ice and dust in the innermost coma of comet 103P/Hartley 2. *Icarus* **238**, 191–204 (2014).
- Brin, G. D. & Mendis, D. A. Dust release and mantle development in comets. *Astrophys. J.* **229**, 402–408 (1979).
- Rickman, H., Fernández, J. A. & Gustafson, B. Å. S. Formation of stable dust mantles on short-period comet nuclei. *Astron. Astrophys.* **237**, 524–535 (1990).
- Jewitt, D. C. From Kuiper belt object to cometary nucleus: the missing ultrared matter. *Astron. J.* **123**, 1039–1049 (2002).

11. Brownlee, D. E. in *Treatise on Geochemistry* (eds Heinrich, D. H. & Karl, K. T.) 663–685 (Pergamon, 2007).
12. Nesvorný, D. *et al.* Cometary origin of the zodiacal cloud and carbonaceous micrometeorites. Implications for hot debris disks. *Astrophys. J.* **713**, 816–836 (2010).
13. Dartois, E. *et al.* Ultracarbonaceous Antarctic micrometeorites, probing the Solar System beyond the nitrogen snow-line. *Icarus* **224**, 243–252 (2013).
14. Guilbert-Lepoutre, A. *et al.* Pre-perihelion activity of comet 67P/Churyumov-Gerasimenko. *Astron. Astrophys.* **567**, <http://dx.doi.org/10.1051/0004-6361/201424186> (2014).
15. Trigo-Rodríguez, J. M. & Llorca, J. On the sodium overabundance in cometary meteoroids. *Adv. Space Res.* **39**, 517–525 (2007).
16. Schulz, R., Alexander, C., Boehnhardt, H. & Glassmeier, K.-H. *Rosetta—ESA's Mission to the Origin of the Solar System* (Springer, 2009).
17. Schulz, R. Rosetta—one comet rendezvous and two asteroid fly-bys. *Sol. Syst. Res.* **43**, 343–352 (2009).
18. Kissel, J. *et al.* COSIMA—high resolution time-of-flight secondary ion mass spectrometer for the analysis of cometary dust particles onboard ROSETTA. *Space Sci. Rev.* **128**, 823–867 (2007).
19. Hornung, K. *et al.* Collecting cometary dust particles on metal blacks with the COSIMA instrument onboard ROSETTA. *Planet. Space Sci.* **103**, 309–317 (2014).
20. DiSanti, M. A. *et al.* Temporal evolution of parent volatiles and dust in Comet 9P/Tempel 1 resulting from the Deep Impact experiment. *Icarus* **187**, 240–252 (2007).
21. Villanueva, G. L. *et al.* The molecular composition of comet C/2007 W1 (Boattini): evidence of a peculiar outgassing and a rich chemistry. *Icarus* **216**, 227–240 (2011).
22. Tozzi, G. P. *et al.* Sublimating components in the coma of comet C/2000 WM1 (LINEAR). *Astron. Astrophys.* **424**, 325–330 (2004).
23. Tsou, P. *et al.* Stardust encounters comet 81P/Wild 2. *J. Geophys. Res.* **109**, E12S01 (2004).
24. Brownlee, D. The Stardust Mission: analyzing samples from the edge of the Solar System. *Annu. Rev. Earth Planet. Sci.* **42**, 179–205 (2014).
25. Kearsley, A. T. *et al.* Experimental impact features in Stardust aerogel: how track morphology reflects particle structure, composition, and density. *Meteorit. Planet. Sci.* **47**, 737–762 (2012).
26. Krueger, H. *et al.* COSIMA-Rosetta calibration for in-situ characterization of 67P/Churyumov-Gerasimenko cometary inorganic compounds. *Planet. Space Sci.* (submitted); preprint at <http://arxiv.org/abs/1501.00716> (2015).
27. Stephan, T. Assessing the elemental composition of comet 81P/Wild 2 by analyzing dust collected by Stardust. *Space Sci. Rev.* **138**, 247–258 (2008).
28. Jessberger, E. K., Christoforidis, A. & Kissel, J. Aspects of the major element composition of Halley's dust. *Nature* **332**, 691–695 (1988).
29. Lodders, K. Solar System abundances and condensation temperatures of the elements. *Astrophys. J.* **591**, 1220–1247 (2003).
30. Jewitt, D. C. & Meech, K. J. Surface brightness profiles of 10 comets. *Astrophys. J.* **317**, 992–1001 (1987).

Acknowledgements COSIMA was built by a consortium led by the Max-Planck-Institut für Extraterrestrische Physik, Garching, Germany, in collaboration with the Laboratoire de Physique et Chimie de l'Environnement et de l'Espace, Orléans, France, the Institut d'Astrophysique Spatiale, CNRS/Université Paris Sud, Orsay, France, the Finnish Meteorological Institute, Helsinki, Finland, the Universität Wuppertal, Wuppertal, Germany, von Hoerner und Sulger GmbH, Schwetzingen, Germany, the Universität der Bundeswehr, Neubiberg, Germany, the Institut für Physik, Forschungszentrum Seibersdorf, Seibersdorf, Austria, the Institut für Weltraumforschung, Österreichische Akademie der Wissenschaften, Graz, Austria and is led by the Max-Planck-Institut für Sonnensystemforschung, Göttingen, Germany. The support of the national funding agencies of Germany (DLR, grant 50 QP 1302), France (CNES), Austria, Finland and the ESA Technical Directorate is gratefully acknowledged. S.S. acknowledges the support by the Swedish National Space Board grant (contract number 121/11). We thank the Rosetta Science Ground Segment at the European Space Astronomy Centre, the Rosetta Mission Operations Centre at the European Space Operations Centre, and the Rosetta Project at the European Space Research and Technology Centre for their work, which enabled the science return of the Rosetta mission.

Author Contributions M.H. managed the project. J.K., Y.L., J.S., K.H., L.T., J.R., K.V. and R.S. contributed to instrument development. J.R., M.H., H.F., Y.L., J.P., L.T., and O.S. contributed to instrument operations and data distribution. C.B., C.E., D.B., A.B., H.C., N.F., M.G., J.K., K.H., H.L., Y.L., L.L.R., F.-R.O.-D., S.M., J.R., J.S., S.S., L.T., B.Z. and M.H. contributed to instrument and data calibration. Y.L. provided grain images and the porosity value. C.E. provided calibrated mass spectrometry data. R.S. performed comet research and wrote the manuscript. All authors discussed the results and commented on the manuscript.

Author Information After the proprietary period of six months the data will be available in the ESA Planetary Science Archive (<http://www.rssd.esa.int/index.php?project=PSA>). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.S. (rita.schulz@esa.int).

Boron isotope evidence for oceanic carbon dioxide leakage during the last deglaciation

M. A. Martínez-Botí^{1*}, G. Marino^{2,3*}, G. L. Foster¹, P. Ziveri^{2,4,5}, M. J. Henehan^{1,6}, J. W. B. Rae^{7,8}, P. G. Mortyn^{2,9} & D. Vance¹⁰

Atmospheric CO₂ fluctuations over glacial–interglacial cycles remain a major challenge to our understanding of the carbon cycle and the climate system. Leading hypotheses put forward to explain glacial–interglacial atmospheric CO₂ variations invoke changes in deep-ocean carbon storage^{1,2}, probably modulated by processes in the Southern Ocean, where much of the deep ocean is ventilated³. A central aspect of such models is that, during deglaciations, an isolated glacial deep-ocean carbon reservoir is reconnected with the atmosphere, driving the atmospheric CO₂ rise observed in ice-core records^{4–6}. However, direct documentation of changes in surface ocean carbon content and the associated transfer of carbon to the atmosphere during deglaciations has been hindered by the lack of proxy reconstructions that unambiguously reflect the oceanic carbonate system. Radiocarbon activity tracks changes in ocean ventilation⁶, but not in ocean carbon content, whereas proxies that record increased deglacial upwelling^{4,7} do not constrain the proportion of upwelled carbon that is degassed relative to that which is taken up by the biological pump. Here we apply the boron isotope pH proxy in planktic foraminifera to two sediment cores from the sub-Antarctic Atlantic and the eastern equatorial Pacific as a more direct tracer of oceanic CO₂ outgassing. We show that surface waters at both locations, which partly derive from deep water upwelled in the Southern Ocean^{8,9}, became a significant source of carbon to the atmosphere during the last deglaciation, when the concentration of atmospheric CO₂ was increasing. This oceanic CO₂ outgassing supports the view that the ventilation of a deep-ocean carbon reservoir in the Southern Ocean had a key role in the deglacial CO₂ rise, although our results allow for the possibility that processes operating in other regions may also have been important for the glacial–interglacial ocean–atmosphere exchange of carbon.

The modern Southern Ocean is a region of vigorous upwelling of carbon- and nutrient-rich waters³. Much of the upwelled CO₂ is outgassed to the atmosphere, owing to incomplete nutrient utilization in the Southern Ocean surface³. These waters are then resubducted as intermediate waters and feed the thermocline of the low-latitude oceans, such as the eastern equatorial Pacific Ocean¹⁰ (EEP), which is at present one of the main oceanic sources of CO₂ to the atmosphere¹¹ (Fig. 1). Many of the mechanisms proposed to reduce atmospheric CO₂ during glacial periods focus on a reduction of Southern Ocean CO₂ leakage, via increased ocean stratification¹² or a more efficient biological pump (probably boosted by iron fertilization)¹³, or both. During deglaciation, this situation is reversed, and previously isolated deep-ocean carbon is thought to be upwelled and re-exposed to the atmosphere⁴.

Deglacial upwelling of aged, nutrient-rich waters in the Southern Ocean, and their subsequent advection to the EEP, has been suggested on the basis of several proxies^{4,6,7}, including biogenic opal fluxes, stable carbon isotope ratios ($\delta^{13}\text{C}$) and radiocarbon activities ($\Delta^{14}\text{C}$). Although these reconstructions provide valuable insights into ocean nutrient dynamics, circulation, and ventilation history, none of them directly tracks ocean–atmosphere CO₂ exchange. For instance, if biological productivity efficiently used upwelled nutrients and carbon, as possibly indicated by the deglacial increase in Southern Ocean and EEP opal fluxes^{4,14}, then CO₂ leakage to the atmosphere may have been damped or even negated completely. Similarly, the appearance of low $\delta^{13}\text{C}$ signatures in the upper ocean at a wide number of locations globally during the deglaciation^{7,15}, which is usually taken as evidence for the re-ventilation of nutrient-enriched (low- $\delta^{13}\text{C}$) waters, could be modulated by changing air–sea CO₂ fractionation¹⁵. Finally, perhaps the most compelling evidence so far for the importance of the recommunication of a deep-ocean

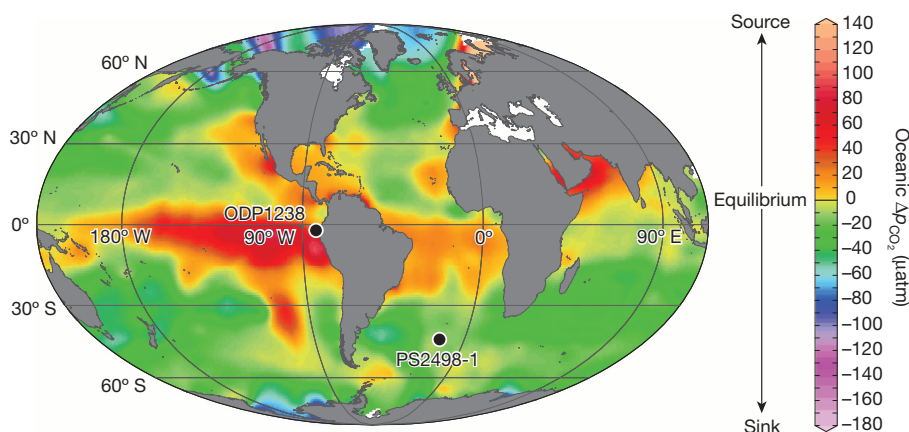


Figure 1 | Location of cores PS2498-1 and ODP1238. Site locations are overlaid on a map of mean annual $\Delta p\text{CO}_2$ (ref. 11).

¹Ocean and Earth Science, National Oceanography Centre Southampton, University of Southampton Waterfront Campus, Southampton SO14 3ZH, UK. ²Institute of Environmental Science and Technology (ICTA), Universitat Autònoma de Barcelona, Bellaterra, Catalonia, 08193, Spain. ³Research School of Earth Sciences, The Australian National University, Canberra, Australian Capital Territory 2601, Australia. ⁴Institució Catalana de Recerca i Estudis Avançats, ICREA, Barcelona, Catalonia, 08010, Spain. ⁵Earth and Climate Cluster, Department of Earth Sciences, Faculty of Earth and Life Sciences, VU Universiteit Amsterdam, 1081HV Amsterdam, The Netherlands. ⁶Department of Geology and Geophysics, Yale University, New Haven, Connecticut 06511, USA. ⁷Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, California 91125, USA. ⁸Department of Earth and Environmental Sciences, University of St Andrews, St Andrews KY16 9AL, UK. ⁹Department of Geography, Universitat Autònoma de Barcelona, Bellaterra, Catalonia, 08193, Spain. ¹⁰Institute of Geochemistry and Petrology, Department of Earth Sciences, ETH Zürich, NW D81.4, Zürich 8092, Switzerland.

*These authors contributed equally to this work.

carbon reservoir with the atmosphere via the Southern Ocean comes from the collapse in $\Delta^{14}\text{C}$ gradients in deep waters around Antarctica due to the deglacial breakdown of stratification in that region⁶. However, although this correlates well with the atmospheric $\Delta^{14}\text{C}$ record and the first rise in atmospheric partial pressure of CO_2 ($p_{\text{CO}_2}^{\text{atm}}$) during Heinrich stadial 1, there is little response during the second rise of $p_{\text{CO}_2}^{\text{atm}}$ centred on the Younger Dryas stadial.

There is therefore an urgent requirement for direct evidence of deglacial CO_2 variations in the surface waters of key areas of the global ocean, such as the Southern Ocean and the EEP, using which it will be possible to test the hypothesis that oceanic CO_2 outgassing was central to the glacial–interglacial $p_{\text{CO}_2}^{\text{atm}}$ rise^{4,7}. We addressed this issue by analysing the boron isotopic composition ($\delta^{11}\text{B}$) of planktic foraminifera, a proxy^{16,17} for oceanic pH that provides a direct link to seawater CO_2 content. Figure 2 compares records of $p_{\text{CO}_2}^{\text{atm}}$ (ref. 5) with new planktic foraminiferal $\delta^{11}\text{B}$ and Mg/Ca-derived temperature for the sub-Antarctic Atlantic (SAA) (site PS2498-1; 44.15° S, 14.23° W, 3,783 m water depth) and the EEP (site ODP1238; 1.87° S, 82.78° W, 2,203 m water depth) (Fig. 1). Surface waters at PS2498-1 are at present influenced by water upwelled in the Antarctic zone and advected northwards via Ekman pumping⁸, whereas ODP1238 is mostly influenced by the Pacific equatorial undercurrent¹⁸, a subsurface current originating in the western equatorial Pacific and fed by water masses of Southern Ocean (~70%) and North Pacific (~30%) origin¹⁹. Previous literature has suggested an effective connection between the sub-Antarctic zone and the EEP via intermediate waters¹⁸, through a process often referred to as ‘oceanic tunnelling’ (Methods).

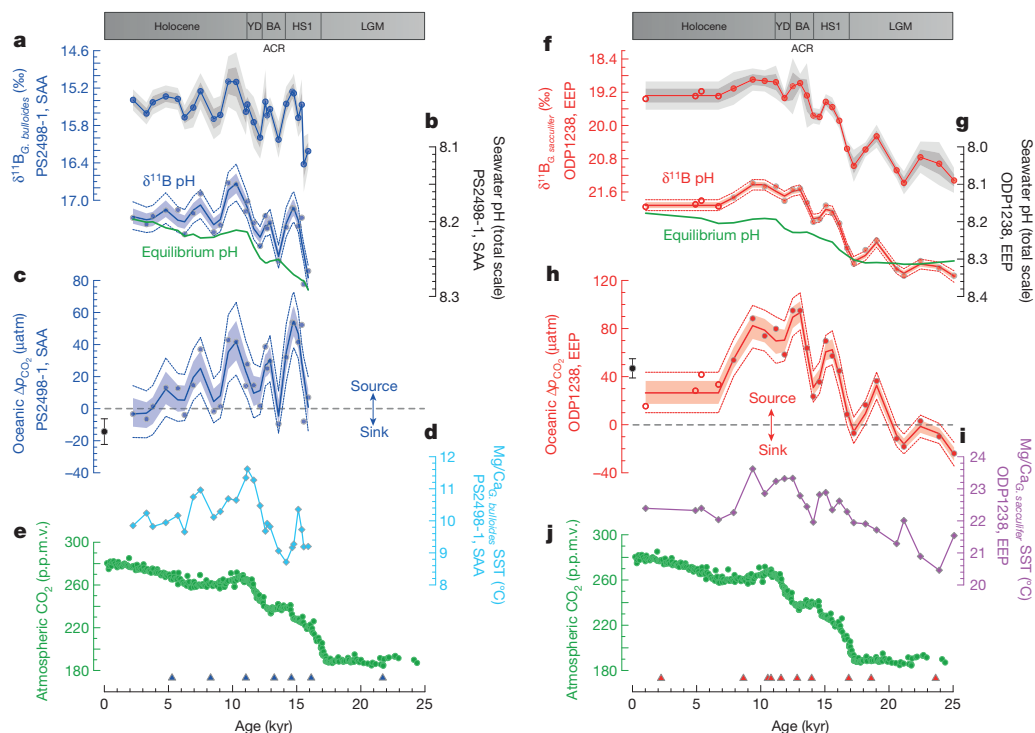


Figure 2 | $\delta^{11}\text{B}$, pH and Δp_{CO_2} records from the SAA and the EEP during the last deglaciation. **a–e**, SAA; **f–j**, EEP. **a**, *Globigerina bulloides* $\delta^{11}\text{B} = ((^{11}\text{B}/^{10}\text{B})_{\text{sample}} / (^{11}\text{B}/^{10}\text{B})_{\text{NIST951}} - 1) \times 1,000\text{‰}$ (Methods) with analytical uncertainties (1 σ , dark grey envelope; 2 σ , light grey envelope). **b**, $\delta^{11}\text{B}$ -based pH (blue) reconstruction for PS2498-1 and calculated seawater pH in equilibrium with atmospheric p_{CO_2} ($p_{\text{CO}_2}^{\text{atm}}$) (green curve) at the same site. **c**, $\delta^{11}\text{B}$ -derived Δp_{CO_2} ($p_{\text{CO}_2}^{\text{sw}} - p_{\text{CO}_2}^{\text{atm}}$). Filled black circle denotes present-day annual average Δp_{CO_2} near PS2498-1 with 2 σ uncertainties¹¹. **d**, *Globigerina bulloides* Mg/Ca-based sea surface temperature (SST). **e**, $p_{\text{CO}_2}^{\text{atm}}$ from Antarctic ice cores⁵. **f**, *Globigerinoides sacculifer* $\delta^{11}\text{B}$ with analytical uncertainties (1 σ , dark grey shading; 2 σ , light grey shading). **g**, $\delta^{11}\text{B}$ -based pH (red) for ODP1238 and calculated seawater pH in equilibrium with $p_{\text{CO}_2}^{\text{atm}}$ (ref. 5) (green curve).

The $\delta^{11}\text{B}$ record for the surface-dwelling foraminifer *Globigerina bulloides* in the sub-Antarctic core PS2498-1 displays a change of ~1.3‰ between the early deglaciation and the Holocene epoch (~0.13 pH units; Fig. 2a, b). Unfortunately, the very low abundance of *G. bulloides* before ~16 kyr ago precluded $\delta^{11}\text{B}$ analyses for the Last Glacial Maximum, which limited our evaluation of the full glacial–interglacial $\delta^{11}\text{B}$ (and pH) shift at this site. The PS2498-1 $\delta^{11}\text{B}$ record does not exhibit a gradual change through the deglaciation into the Holocene, but features two distinct decreases (lower pH). In the EEP (at ODP1238), there is a brief negative excursion in the $\delta^{11}\text{B}$ of the surface-dweller *Globigerinoides sacculifer* at 19 kyr ago, and the ensuing deglacial change is larger (~2‰; ~0.2 pH units) and more gradual than that recorded by *G. bulloides* at PS2498-1 (Fig. 2f, g and Extended Data Fig. 1).

To highlight the main patterns of variability in the $\delta^{11}\text{B}$ -derived time series and to probabilistically account for all the uncertainties associated with our reconstructions, we used a Monte Carlo approach that uses a non-parametric regression (LOESS function; Fig. 2b, c, g, h and Methods). A comparison of the PS2498-1 and ODP1238 pH records with the pH expected if surface waters remained in equilibrium with the atmosphere (Fig. 2b, g, green lines), reveals ‘excess’ surface ocean acidification in both these areas during the deglaciation and early Holocene. To gain further insight into deglacial ocean–atmosphere CO_2 exchange, we use the $\delta^{11}\text{B}$ -derived pH data, along with estimates of temperature, salinity and alkalinity, to calculate the partial pressure of CO_2 in seawater²⁰ ($p_{\text{CO}_2}^{\text{sw}}$; Methods). Note that the temperature, salinity and alkalinity have a minor effect on $p_{\text{CO}_2}^{\text{sw}}$ (at most ± 13 μatm , ± 15 μatm and ± 12 μatm , respectively), which is driven mainly by the $\delta^{11}\text{B}$ -derived

h, $\delta^{11}\text{B}$ -based Δp_{CO_2} . Filled black circle denotes present-day mean annual Δp_{CO_2} near ODP1238 with 2 σ uncertainties¹¹. **i**, *Globigerinoides sacculifer* Mg/Ca-based SST. The late-Holocene data in **f**, **g** and **h** have been averaged (individual measurements are shown as open red circles) (Methods). Envelopes in **b**, **c**, **g** and **h** are 68% and 95% uncertainty bounds (light blue or red shading and dotted lines, respectively) based on a LOESS regression of the $\delta^{11}\text{B}$ -derived records using a Monte Carlo approach; thick line denotes the maximum-probability fit to the data. YD, Younger Dryas; BA, Bølling–Allerød; ACR, Antarctic cold reversal; HS1, Heinrich stadial 1; LGM, Last Glacial Maximum. Filled triangles at the bottom indicate calibrated ^{14}C ages (Methods).

pH, given the strong correlation of pH and $[\text{CO}_2]$ within the ocean carbonate system²⁰. The $p_{\text{CO}_2}^{\text{sw}}$ profiles are compared with $p_{\text{CO}_2}^{\text{atm}}$ from ice-core records to yield $\Delta p_{\text{CO}_2} = p_{\text{CO}_2}^{\text{sw}} - p_{\text{CO}_2}^{\text{atm}}$, and, notably, our late-Holocene Δp_{CO_2} estimates agree within uncertainties with modern water-column data from nearby locations¹¹, supporting the accuracy of our $\delta^{11}\text{B}$ -pH calibrations (Fig. 2c–h).

Our EEP Δp_{CO_2} record reveals that this region became a significant source of CO_2 to the atmosphere during the deglaciation and the early Holocene, reaching, ~ 13 kyr ago, a peak of $+90 \pm 16 \mu\text{atm}$ (2σ), which exceeds by $\sim 45 \mu\text{atm}$ the average modern values at this location¹¹ (Fig. 2h). Comparison with earlier studies in the western and central Pacific^{21–23} that reported deglacial Δp_{CO_2} values of $+100$ to $\geq +185 \mu\text{atm}$ shows the widespread nature of the CO_2 outgassing in the equatorial Pacific (Extended Data Fig. 2). The Δp_{CO_2} data from the SAA also indicate that this region acted as a source of CO_2 to the atmosphere from ~ 16 to ~ 7.5 kyr ago, but in the form of two prominent multimillennial events, with a maximum of $+50 \pm 18 \mu\text{atm}$ (that is, $\sim 65 \mu\text{atm}$ higher than present-day average values¹¹) ~ 15 kyr ago (Fig. 2c).

The occurrence of upper-ocean acidification in the SAA and the EEP during the deglaciation coincided, within uncertainties, with excursions to higher opal fluxes at nearby sites^{4,14} (Fig. 3a, b, f, g). This observation provides compelling evidence for the proposed link between, on the one hand, the resumption of Antarctic upwelling⁴ and attendant Ekman transport^{8,10} of deep waters enriched in nutrients and $[\text{CO}_2]$, and, on the other hand, the increase in oceanic Δp_{CO_2} in the wider Southern Ocean. It further suggests that advection to the EEP, via oceanic tunnelling, of these CO_2 -rich waters may (at least partly) explain the deglacial Δp_{CO_2} maximum documented at ODP1238. This interpretation is corroborated by the broad synchronicity of the acidification phases with excursions to depleted surface ocean $\delta^{13}\text{C}$ at our core locations (Fig. 3c, h), with the deglacial rise in the concentration of atmospheric CO_2 (Fig. 3e, l), and with the negative shift⁵ in the $\delta^{13}\text{C}$ of atmospheric CO_2 (Fig. 3e, k).

However, the differences in the patterns of pH, Δp_{CO_2} , $\delta^{13}\text{C}$ and opal at the two locations (Figs 2 and 3 and Extended Data Fig. 1), suggest that the EEP record does not represent merely a downstream expression of the deglacial CO_2 outgassing in the Southern Ocean, and that additional processes or CO_2 sources, or both, need to be considered to fully explain its structure. Intermediate waters originating in the Southern Ocean incorporate remineralized carbon during transit to the EEP, which may modify their geochemical characteristics²⁴. Additionally, because source waters of the Pacific equatorial undercurrent also include waters originating in the North Pacific¹⁹ ($\sim 30\%$), these may provide a potential supplementary source²⁵ of CO_2 . During the early Holocene, both our Δp_{CO_2} records also indicate continued oceanic CO_2 outgassing, yet $p_{\text{CO}_2}^{\text{atm}}$ shows little change. This suggests that ocean outgassing at this time was balanced by CO_2 uptake, possibly due to forest regrowth and peat build-up as the continental ice sheets retreated from the northern continents^{2,26}, in line with the near-contemporaneous rise⁵ in atmospheric $\delta^{13}\text{C}$ (Fig. 3e, k).

Although millennial changes in the mean position of Southern Ocean frontal systems²⁷, and in the upwelling strength in the EEP due to shifts in the intertropical convergence zone²⁸, have the potential to change Δp_{CO_2} at our sites, such changes seem unlikely to exert a dominant influence on our records. If the high Δp_{CO_2} spikes we document were driven by a northward shift of the regional oceanic fronts (PS2498-1) or by increased upwelling (ODP1238), they should be accompanied by surface cooling, whereas in fact both sites feature a warming trend during these intervals (Fig. 2 and Methods). It is also unlikely that the signals are driven by migration in foraminiferal depth habitat. In the SAA, pH vertical gradients are small, whereas in the EEP migration to deeper and more-acidic waters is inconsistent with the warming signal in Mg/Ca (Fig. 2) and with the structural similarity between *G. sacculifer* $\delta^{18}\text{O}$ and $\delta^{13}\text{C}$ records and those for the strict surface-dweller *Globigerinoides ruber* (Methods and Extended Data Fig. 3).

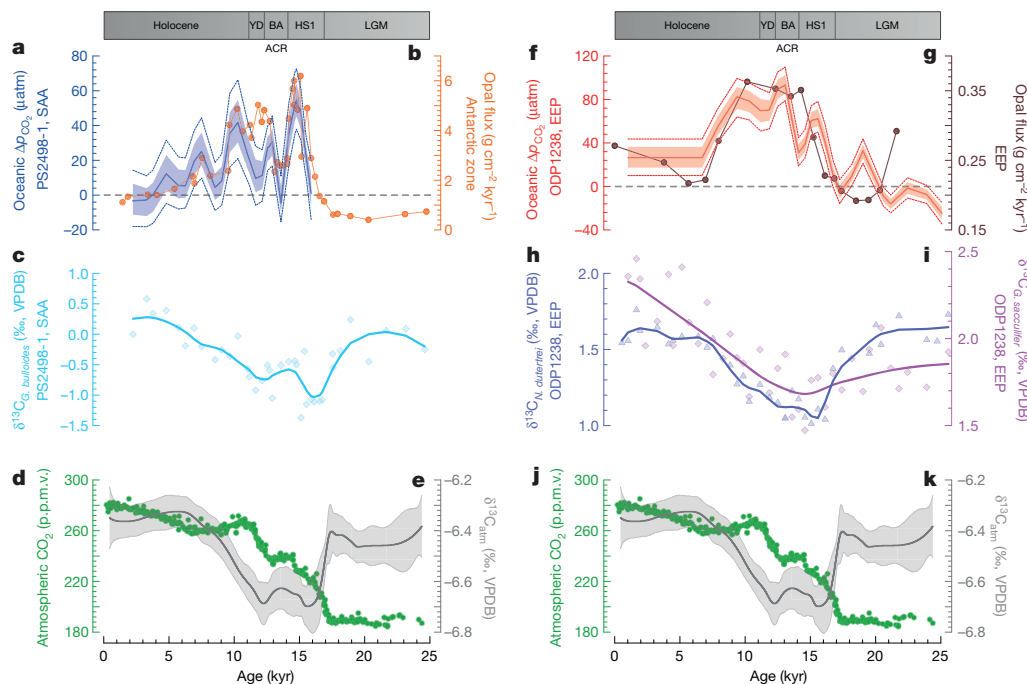


Figure 3 | Δp_{CO_2} and $\delta^{13}\text{C}$ records from the SAA and the EEP during the last deglaciation compared to opal fluxes, $p_{\text{CO}_2}^{\text{atm}}$ and atmospheric $\delta^{13}\text{C}$. **a–e**, SAA; **f–k**, EEP. **a**, Surface ocean Δp_{CO_2} reconstruction for PS2498-1. **b**, Opal fluxes for TN057-13-4PC (Antarctic zone)⁴. **c**, PS2498-1 *G. bulloides* $\delta^{13}\text{C} = ((^{13}\text{C}/^{12}\text{C})_{\text{sample}} / (^{13}\text{C}/^{12}\text{C})_{\text{VPDB}} - 1) \times 1,000\text{‰}$ (blue dots; VPDB, Vienna Pee Dee Belemnite) with a LOESS non-parametric regression of the data (span = 0.4) (solid line; Methods). **d**, **j**, Atmospheric CO_2 concentrations from Antarctic ice cores⁵. **e**, **k**, Atmospheric $\delta^{13}\text{C}$ record from Antarctic ice cores⁵. **f**, Surface ocean Δp_{CO_2} reconstruction in ODP1238. **g**, Opal fluxes

for V19-30 (EEP)¹⁴. **h**, **i**, ODP1238 *G. sacculifer* (**i**) and *Neogloboquadrina dutertrei* (**h**) $\delta^{13}\text{C}$ with a LOESS non-parametric regression of the data (solid lines; span = 0.7 (**h**) and 0.33 (**i**)). Envelopes in **a** and **f** are 68% and 95% uncertainty bounds (light blue or red shading and dotted lines, respectively) based on a LOESS regression of the $\delta^{11}\text{B}$ -derived records using a Monte Carlo approach; thick line denotes the maximum-probability fit to the data. Envelopes in **e** and **k** are 2σ uncertainty bounds around the Monte Carlo average (thick line)⁵.

Local changes in the efficiency of the biological pump, stimulated by iron fertilization, also have the potential¹ to influence surface water $p\text{CO}_2$. The evidence that the EEP was a weaker CO_2 source (or even a sink) during the Last Glacial Maximum may be ascribed to a dust-derived relaxation of iron limitation and attendant strengthening of the biological pump^{29,30}. It is also possible that the gradual decrease in the iron supply to the EEP and sub-Antarctic regions during the deglaciation^{13,29} could explain part of the high deglacial $\Delta p\text{CO}_2$ at these sites. However, the marked differences between the structure of the dust flux records and our $\Delta p\text{CO}_2$ reconstructions argue against this hypothesis (Extended Data Fig. 4).

It therefore seems most plausible that the $\Delta p\text{CO}_2$ pulses that we report here using planktic foraminifera $\delta^{11}\text{B}$ reflect oceanic CO_2 outgassing during the last deglaciation. A common origin for the $\Delta p\text{CO}_2$ anomalies presented here probably involved the renewed upwelling of aged⁶ deep water enriched in nutrients⁴ and carbon in the Southern Ocean, which subsequently 'leaked' into the atmosphere, contributing to the deglacial $p\text{CO}_2^{\text{atm}}$ rise and the atmospheric $\Delta^{14}\text{C}$ and $\delta^{13}\text{C}$ decreases^{5,31}. However, our analysis also indicates that other carbon sources, possibly located in the wider Pacific Ocean²⁵, may need to be invoked to fully explain glacial–interglacial ocean–atmosphere CO_2 exchange.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 22 August; accepted 11 December 2014.

- Köhler, P., Fischer, H., Munhoven, G. & Zeebe, R. E. Quantitative interpretation of atmospheric carbon records over the last glacial termination. *Glob. Biogeochem. Cycles* **19**, GB4020 (2005).
- Yu, J. *et al.* Loss of carbon from the deep sea since the Last Glacial Maximum. *Science* **330**, 1084–1087 (2010).
- Sigman, D. M., Hain, M. P. & Haug, G. H. The polar ocean and glacial cycles in atmospheric CO_2 concentration. *Nature* **466**, 47–55 (2010).
- Anderson, R. F. *et al.* Wind-driven upwelling in the Southern Ocean and the deglacial rise in atmospheric CO_2 . *Science* **323**, 1443–1448 (2009).
- Schmitt, J. *et al.* Carbon isotope constraints on the deglacial CO_2 rise from ice cores. *Science* **336**, 711–714 (2012).
- Burke, A. & Robinson, L. F. The Southern Ocean's role in carbon exchange during the last deglaciation. *Science* **335**, 557–561 (2012).
- Spero, H. J. & Lea, D. W. The cause of carbon isotope minimum events on glacial terminations. *Science* **296**, 522–525 (2002).
- Marshall, J. & Speer, K. Closure of the meridional overturning circulation through Southern Ocean upwelling. *Nature Geosci.* **5**, 171–180 (2012).
- Liu, Z. Y. & Yang, H. J. Extratropical control of tropical climate, the atmospheric bridge and oceanic tunnel. *Geophys. Res. Lett.* **30**, 1230 (2003).
- Sarmiento, J. L., Gruber, N., Brzezinski, M. A. & Dunne, J. P. High-latitude controls of thermocline nutrients and low-latitude biological productivity. *Nature* **427**, 56–60 (2004).
- Takahashi, T. *et al.* Climatological mean and decadal change in surface ocean $p\text{CO}_2$, and net sea-air CO_2 flux over the global oceans. *Deep-Sea Res.* **56**, 2075–2076 (2009).
- Keeling, R. F. & Visbeck, M. Palaeoceanography: Antarctic stratification and glacial CO_2 . *Nature* **412**, 605–606 (2001).
- Martínez-García, A. *et al.* Links between iron supply, marine productivity, sea surface temperature, and CO_2 over the last 1.1 Ma. *Paleoceanography* **24**, PA1207 (2009).
- Bradt Miller, L. I., Anderson, R. F., Fleisher, M. Q. & Burckle, L. H. Diatom productivity in the equatorial Pacific Ocean from the last glacial period to the present: a test of the silicic acid leakage hypothesis. *Paleoceanography* **21**, PA4201 (2006).
- Ninnemann, U. S. & Charles, C. D. Regional differences in Quaternary Subantarctic nutrient cycling: Link to intermediate and deep water ventilation. *Paleoceanography* **12**, 560–567 (1997).
- Hönisch, B. & Hemming, N. G. Surface ocean pH response to variations in $p\text{CO}_2$ through two full glacial cycles. *Earth Planet. Sci. Lett.* **236**, 305–314 (2005).
- Foster, G. L. Seawater pH, $p\text{CO}_2$ and $[\text{CO}_2^{-3}]$ variations in the Caribbean Sea over the last 130 kyr: a boron isotope and B/Ca study of planktic foraminifera. *Earth Planet. Sci. Lett.* **271**, 254–266 (2008).
- Toggweiler, J. R., Dixon, K. & Broecker, W. S. The Peru upwelling and the ventilation of the South Pacific thermocline. *J. Geophys. Res.* **96**, 20467–20497 (1991).
- Goodman, P. J., Hazeleger, W., De Vries, P. & Cane, M. Pathways into the Pacific equatorial undercurrent: a trajectory analysis. *J. Phys. Oceanogr.* **35**, 2134–2151 (2005).
- Zeebe, R. E. & Wolf-Gladrow, D. A. *CO_2 in Seawater: Equilibrium, Kinetics, Isotopes* (Elsevier, 2001).
- Palmer, M. R. & Pearson, P. N. A 23,000-year record of surface water pH and $p\text{CO}_2$ in the western equatorial Pacific Ocean. *Science* **300**, 480–482 (2003).
- Kubota, K., Yokoyama, Y., Ishikawa, T., Obrochta, S. & Suzuki, A. Larger CO_2 source at the equatorial Pacific during the last deglaciation. *Sci. Rep.* **4**, 5261 (2014).
- Douville, E. *et al.* Abrupt sea surface pH change at the end of the Younger Dryas in the central sub-equatorial Pacific inferred from boron isotope abundance in corals (*Porites*). *Biogeosciences* **7**, 2445–2459 (2010).
- Palter, J. B., Sarmiento, J. L., Gnanadesikan, A., Simeon, J. & Slater, R. D. Fueling export production: nutrient return pathways from the deep ocean and their dependence on the meridional overturning circulation. *Biogeosciences* **7**, 3549–3568 (2010).
- Rae, J. W. B. *et al.* Deep water formation in the North Pacific and deglacial CO_2 rise. *Paleoceanography* **29**, 645–667 (2014).
- Menviel, L. & Joos, F. Toward explaining the Holocene carbon dioxide and carbon isotope records: results from transient ocean carbon cycle–climate simulations. *Paleoceanography* **27**, PA1207 (2012).
- Barker, S. *et al.* Interhemispheric Atlantic seesaw response during the last deglaciation. *Nature* **457**, 1097–1102 (2009).
- Denton, G. H. *et al.* The last glacial termination. *Science* **328**, 1652–1656 (2010).
- McGee, D., Marcantonio, F. & Lynch-Stieglitz, J. Deglacial changes in dust flux in the eastern equatorial Pacific. *Earth Planet. Sci. Lett.* **257**, 215–230 (2007).
- Pichevin, L. E. *et al.* Enhanced carbon pump inferred from relaxation of nutrient limitation in the glacial ocean. *Nature* **459**, 1114–1117 (2009).
- Reimer, P. J. *et al.* IntCal13 and Marine13 radiocarbon age calibration curves 0–50,000 years cal BP. *Radiocarbon* **55**, 1869–1887 (2013).

Acknowledgements We thank the International Ocean Drilling Program for providing samples from ODP Leg 202, R. Gersonde and A. Mackensen for the PS2498-1 core material, J. F. McManus for sharing his unpublished benthic isotope data for ODP1238, and E. J. Rohling, M. P. Hain and C. Beaulieu for discussions. For the calibration of *G. bulloides*, we thank M. Kucera for providing core-top samples from the archives at the University of Tübingen, H. C. Bostock for samples from the National Institute for Water and Atmospheric Research, Wellington, and B. J. Marshall and R. Thunell for samples from the Cariaco Basin sediment trap time series. J. A. Milton, M. J. Cooper and A. Michalik provided assistance during ICP-MS analyses and sample preparation in the laboratory. C. Alt and M. T. Horigome helped with foraminifera picking. We thank the other members of 'The B-Team' at the National Oceanography Centre Southampton for their contributions. Financial support was provided by the European Community through a Marie Curie Intra-European Fellowship for Career Development to M.A.M.-B., the Universitat Autònoma de Barcelona through a Postdoctoral Research Grant to G.M., the Spanish Ministry of Science and Innovation (PROCARSO project CGL2009-10806) to G.M., P.Z. and P.G.M., a NERC PhD studentship awarded to M.J.H., a NOAA/UCAR Climate and Global Change Postdoctoral Fellowship to J.W.B.R., and NERC grant NE/D00876/X2 to G.L.F. G.M. was also supported by the Australian Laureate Fellowship project FL120100050 (E. J. Rohling).

Author Contributions M.A.M.-B., G.M., G.L.F. and P.Z. designed the study; G.M. and M.A.M.-B. produced the $\delta^{11}\text{B}$ and trace element records for PS2498-1; M.A.M.-B. produced the $\delta^{11}\text{B}$ and trace element records for ODP1238; G.M. produced the $\delta^{18}\text{O}$ and $\delta^{13}\text{C}$ data; M.J.H. developed the *G. bulloides* $\delta^{11}\text{B}$ –pH calibration; J.W.B.R. sampled sediment core PS2498-1. M.A.M.-B. and G.M. wrote the first draft jointly, and all authors contributed to the interpretation and the preparation of the final manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.A.M.-B. (m.a.martinez-boti@noc.soton.ac.uk) and G.M. (gianluca.marino@anu.edu.au).

METHODS

Material and methods. Core PS2498-1 was retrieved by the RV *Polarstern* from the eastern flank of the mid-Atlantic Ridge (44.15° S, 14.23° W, 3,783 m water depth), in the sub-Antarctic Atlantic Ocean^{32,33} (Fig. 1). Between 400 and 650 individuals of *G. bulloides* were picked from the 300–355 µm size fraction (10 µg per shell with B/Ca ≈ 35 µmol mol⁻¹) from this site. Ocean Drilling Program (ODP) Site 1238 (1.87° S, 82.78° W, 2,203 m water depth) was drilled during Leg 202 by the RV *Joides Resolution* from the southern flank of Carnegie Ridge, off the shore of Ecuador in the EEP³⁴ (Fig. 1). From ODP1238 we picked 15–90 individuals of *G. sacculifer* of mixed morphotypes (see below) from the 425–500 µm size fraction (40 µg per shell with B/Ca ≈ 100 µmol mol⁻¹) for δ¹¹B and trace elements, and 10–20 individuals from the 355–425 µm size fraction for δ¹⁸O and δ¹³C. Twenty individuals of *G. ruber* sensu stricto from the 250–355 µm size fraction were picked for δ¹⁸O and δ¹³C, and 30 individuals of *N. dutertrei* were picked from the 355–500 µm size fraction for trace elements and δ¹⁸O and δ¹³C. Foraminiferal samples were crushed between cleaned glass microscope slides before cleaning and analysis.

Age models. For core PS2498-1, we converted existing³² accelerator mass spectrometry (AMS) ¹⁴C data to calendar ages using the Calib 7.0 program^{35,36} with the Marine13 data set³¹. Local reservoir corrections of ΔR = 300 yr (ref. 37) and ΔR = 900 yr were used for the past 16 kyr and between 16 and 26 kyr ago, respectively (Extended Data Fig. 5a), following recent studies^{6,38} (compare with ref. 39). Different ΔR values may affect the chronology by up to 0.9 kyr for the LGM and early deglacial sections, although these potential changes in age model do not impact our conclusions unduly (Extended Data Fig. 5b).

The age model for ODP1238 is based on ten new AMS ¹⁴C dates (Extended Data Fig. 6a). Monospecific samples of *N. dutertrei* (~300 individuals, ~12 mg) were analysed at the Lawrence Livermore National Laboratory Center for Accelerator Mass Spectrometry. Carbon-14 ages were calibrated using the Calib 7.0 program^{35,36} with the Marine13 data set³¹. In line with previous studies^{40,41} in the EEP, an average ΔR of 72 ± 35 yr was applied to all ¹⁴C data (<http://calib.qub.ac.uk/marine>). To derive an age model for ODP1238, the relationship between sediment depth and calibrated ages was fitted with a third-order polynomial regression (Extended Data Fig. 6b). A potential complication in constructing ¹⁴C-based age models in the EEP is the possible presence of old waters during some intervals of the last deglaciation^{42,43}. However, the gradual increase in age with depth at ODP1238, with no sign of age reversals (compare with refs 42, 43) suggests that our age model is not likely to be influenced by extremely old ¹⁴C, or that the influence, if present, was small. Moreover, there is good agreement between the benthic and planktic δ¹⁸O records in ODP1238 and those of other EEP cores^{7,40,41,44} (Extended Data Fig. 7). Although this does not represent a definitive validation, given that all these sites could potentially be influenced by the same aged water mass, it documents the coherence of our age model with the published literature for this region.

Foraminiferal signal carriers. *Globigerina bulloides* is ubiquitous in the South Atlantic⁴⁵ and its abundance seems to be generally associated with the phytoplankton productivity maxima during austral summer⁴⁶. In the proximity of PS2498-1, *G. bulloides* dwells in the upper ~60 m of the water column, with a distinct maximum between 0 and 25 m (ref. 46).

A number of proxy systems based on *G. sacculifer* in the EEP have been demonstrated to reflect mean annual environmental conditions^{47–49}. In the Panama Basin the depth distribution of *G. sacculifer* is morphotype dependent⁵⁰: *G. sacculifer* without a sac-like final chamber dwells predominantly in the surface mixed layer, whereas *G. sacculifer* with a sac-like final chamber dwells in the thermocline⁵⁰ (25–37 m). Another study⁵¹ showed that close to the coastal upwelling region of the EEP (where ODP1238 is located) the two morphotypes of *G. sacculifer* have similar depth ranges (0–30 m), which also overlap with that of the surface-dweller *G. ruber* (0–25 m). *Globigerinoides sacculifer* is also known to add gametogenic calcite at thermocline depths, which can account for up to 30% of the final test weight⁵². Although these and other⁵³ studies may imply the influence of a thermocline signal on the *G. sacculifer* mixed-morphotype-based reconstructions, δ¹⁸O and δ¹³C data from ODP1238 (Extended Data Fig. 3) indicate a similar depth habitat for *G. sacculifer* and the surface-dweller *G. ruber*^{50,51} (which does not add gametogenic calcite⁵⁴) throughout the past 25 kyr.

δ¹³C and δ¹⁸O analyses. Prior to stable isotope determination, foraminiferal samples were rinsed with methanol, ultrasonicated and then oven-dried at 40 °C. Analyses of *G. bulloides* (PS2498-1, complementing ref. 55), *G. sacculifer* (ODP1238) and *N. dutertrei* (ODP1238) were performed at the Universitat Autònoma de Barcelona, using a Thermo Finnigan MAT253 mass spectrometer coupled to a Kiel IV device for CO₂ sample gas preparation. External reproducibility (1σ) of carbonate standards was better than ±0.05‰ for δ¹⁸O and ±0.03‰ for δ¹³C. ODP1238 *G. ruber* δ¹⁸O and δ¹³C data were measured at MARUM-Center for Marine Environmental Sciences, University of Bremen, using a Thermo Finnigan MAT252 mass spectrometer coupled to a Bremen-type automatic carbonate preparation device.

The precision (1σ) was better than 0.07‰ for δ¹⁸O and 0.05‰ for δ¹³C, on the basis of replicates of an internal limestone standard.

Trace element and δ¹¹B analyses. Foraminiferal samples were oxidatively cleaned^{56–58}, dissolved in ~0.15 M Teflon-distilled HNO₃ and centrifuged. An aliquot (~20 µl; ~7% of the total sample) was taken for trace element analyses, which were performed on a Thermo Scientific Element 2 single-collector ICPMS at the University of Southampton¹⁷. Mg/Ca ratios were converted to calcification temperatures using the calibration of ref. 59 for *G. sacculifer* (ODP1238) and ref. 60 for *G. bulloides* (PS2498-1). The analytical reproducibility for Mg/Ca was ±2.7% (2σ), and the efficiency of the foraminiferal cleaning^{17,58} was verified using Al/Ca ratios (<100 µmol mol⁻¹ in all samples, and typically <60 µmol mol⁻¹). The *N. dutertrei* (355–500 µm) Mg/Ca record was produced at the University of Bristol and at the University of Southampton using the same standard set to ensure comparability. Samples were reductively cleaned^{57,61} and measured on a Thermo Scientific Element 2 single-collector ICPMS following ref. 17. Calcification temperatures were derived from ref. 59.

Boron was separated from the remaining sample using Amberlite IRA-743 boron-specific anion exchange resin¹⁷. δ¹¹B was measured on a Thermo Scientific Neptune multicollector inductively coupled plasma mass spectrometer (MC-ICPMS) at the University of Southampton^{17,58,62}. The external reproducibility of δ¹¹B analyses was calculated following the approach of ref. 58, and is described by the relationship

$$2\sigma = 1.87 \exp^{-20.6[^{11}\text{B}]} + 0.22 \exp^{-0.43[^{11}\text{B}]} \quad (1)$$

where [¹¹B] is the intensity of the ¹¹B signal in volts⁶².

Owing to the low abundance of *G. sacculifer* during the Holocene in ODP1238, the associated 2σ uncertainties (calculated using equation (1)) were relatively large (0.56–0.84‰). The four most recent Holocene samples have been averaged, with a 2σ uncertainty calculated as the mean of the individual uncertainties (from equation (1)) divided by the square root of $n - 1$ ($n = 4$).

The boron isotope pH proxy. The boron isotope pH proxy has been extensively described in previous studies^{16,17,58}. Briefly, boron in seawater exists mainly as two different species, boric acid (B(OH)₃) and borate ion (B(OH)₄⁻), and their relative abundance is pH dependent. There are two isotopes of boron, ¹¹B (~80%) and ¹⁰B (~20%), with a ratio normally expressed as

$$\delta^{11}\text{B} = \left(\frac{^{11}\text{B}/^{10}\text{B}_{\text{sample}}}{^{11}\text{B}/^{10}\text{B}_{\text{NIST951}}} - 1 \right) \times 1,000\text{‰}$$

where ¹¹B/¹⁰B_{NIST951} is the isotopic ratio of the NIST SRM 951 boric acid standard⁶³ (¹¹B/¹⁰B = 4.04367).

There is pronounced isotopic fractionation between the two boron species, with boric acid being enriched⁶⁴ in ¹¹B by 27.2‰. Because the concentration of each species is pH dependent, their isotopic composition also has to change with pH to maintain a constant seawater δ¹¹B. Calibration studies^{58,65,66} have shown that the borate species is predominantly incorporated into foraminiferal CaCO₃, and ocean pH can therefore be calculated from the δ¹¹B of borate as

$$\text{pH} = \text{pK}^*_\text{B} - \log \left(\frac{\delta^{11}\text{B}_{\text{sw}} - \delta^{11}\text{B}_{\text{borate}}}{\delta^{11}\text{B}_{\text{sw}} - (11-10K_\text{B}\delta^{11}\text{B}_{\text{borate}}) - 1,000((11-10K_\text{B}) - 1)} \right) \quad (2)$$

where pK^*_B is the dissociation constant for boric acid at *in situ* temperature, salinity and pressure⁶⁷, δ¹¹B_{sw} is the isotopic composition of seawater⁶⁸ (39.61‰), δ¹¹B_{borate} is the isotopic composition of borate ion, and $11-10K_\text{B}$ is the isotopic fractionation between the two aqueous species of boron in seawater⁶⁴ (1.0272 ± 0.0006).

Calibration of δ¹¹B pH proxy in *G. bulloides*. We combined core-top and sediment trap data to calibrate the δ¹¹B pH proxy in *G. bulloides* over a broad range (~2‰) of δ¹¹B_{borate} (ref. 69). Core-top samples were taken from core archives at the University of Tübingen (Germany) and NIWA (New Zealand) (Extended Data Fig. 8). Carbon-14 dating⁷⁰ (samples from NIWA) and the presence of Rose Bengal-stained living benthic foraminifera (samples from University of Tübingen) confirmed recent ages for the samples used. pH was estimated for core-top sites using surface water oceanographic data^{11,37,71} (following ref. 62), and regional total alkalinity/salinity/temperature relationships⁷². The pre-industrial partial pressure of CO₂ in seawater ($p_{\text{CO}_2}^{\text{sw}}$) at each core-top site was estimated by applying monthly ocean-atmosphere Δ p_{CO_2} interpolated from surrounding sites¹¹ (and corrected for the post-industrial changes in flux⁷³) to a pre-industrial atmospheric p_{CO_2} value from ice-core data. Monthly estimates of pH were then calculated with CO2sys-Matlab⁷⁴, using the constants of refs 67, 75, 76, and temperature and salinity data¹¹. Monthly-resolved *in situ* δ¹¹B_{borate} could then be calculated from pH, temperature and salinity. For each sample site, the average of twelve monthly estimates of δ¹¹B_{borate} was taken as mean annual *in situ* δ¹¹B_{borate}, and two standard deviations of the monthly variability were taken as 2σ uncertainty (representing intraannual variation in δ¹¹B_{borate} at each site). These core-top samples were complemented by sediment trap samples from the Cariaco Basin CAR22(Z), collected in January 2007. pH, temperature and salinity data for the sediment trap site are interpolated from data from December

2006 and February 2007 (<http://www.imars.usf.edu/CAR>). These new calibration data are presented and plotted in Extended Data Fig. 8. We can then calculate $\delta^{11}\text{B}_{\text{borate}}$ from *G. bulloides* $\delta^{11}\text{B}$ ($\delta^{11}\text{B}_{\text{calcite}}$), with associated 2σ uncertainties, as follows:

$$\delta^{11}\text{B}_{\text{borate}} = (\delta^{11}\text{B}_{\text{calcite}} + 3.440 \pm 4.584) / 1.074 \pm 0.252 \quad (3)$$

Size fraction was not found to produce any effect on $\delta^{11}\text{B}$ (ref. 69); in this study we used the 300–355 μm size fraction.

$\delta^{11}\text{B}$ pH proxy in *G. sacculifer*. Conversion of $\delta^{11}\text{B}$ of *G. sacculifer* to pH follows the approach described in ref. 77. Despite analytical biases between MC-ICPMS and N-TIMS analysis⁸⁸, it can be assumed that the pH sensitivity of *G. sacculifer* $\delta^{11}\text{B}$ described by culture calibration⁷⁸ (N-TIMS) is applicable to MC-ICPMS data^{88,79}. The $\delta^{11}\text{B}$ –pH calibration used here therefore incorporates both culture data⁷⁸, which we recalculated following ref. 62, and existing core-top data¹⁷ (Extended Data Fig. 8). To account for the analytical bias between MC-ICPMS and N-TIMS, we corrected the data set of ref. 78 by applying an offset of -3.32% . This is derived from the comparison between core-top *G. sacculifer* measurements by MC-ICPMS¹⁷ and N-TIMS⁸⁰ from adjacent sites⁸⁰. The following equation (with 2σ uncertainties) can be then used to calculate the appropriate $\delta^{11}\text{B}_{\text{borate}}$ from *G. sacculifer* $\delta^{11}\text{B}$ ($\delta^{11}\text{B}_{\text{calcite}}$):

$$\delta^{11}\text{B}_{\text{borate}} = (\delta^{11}\text{B}_{\text{calcite}} - 3.600 \pm 0.722) / 0.834 \pm 0.036 \quad (4)$$

Carbonate system calculations: pH, $p_{\text{CO}_2}^{\text{sw}}$ and Δp_{CO_2} . To calculate pH from $\delta^{11}\text{B}_{\text{borate}}$ (equation (2)), temperature and salinity estimates are required. Temperatures are derived from foraminiferal Mg/Ca (see above) and salinity is calculated taking into account the freshwater loss from the ocean due to the glacial growth of continental ice sheets⁸¹. A Monte Carlo approach was used to generate 10,000 realizations of pH by randomly sampling the relevant input parameters within their given uncertainty bounds (2σ): $\delta^{11}\text{B}$ plus/minus analytical uncertainty (equation (1)) and plus/minus calibration uncertainty (equations (3) and (4)); temperature $\pm 1^\circ\text{C}$; and salinity ± 1 p.s.u. To fully propagate the calibration uncertainty, 10,000 realizations of each of the calibration data points were first made by randomly varying each point within its x and y uncertainties and fitting a separate regression line for each set of $\delta^{11}\text{B}_{\text{borate}}$ and $\delta^{11}\text{B}_{\text{calcite}}$ values. This approach therefore accounts for the covariation in slope and intercept uncertainty in the average $\delta^{11}\text{B}_{\text{borate}} - \delta^{11}\text{B}_{\text{calcite}}$ calibration line shown in Extended Data Fig. 8 (note that this covariation arises because a fit to the data with a high slope has a low intercept, and vice versa). For each of the 10,000 realizations of the down-core $\delta^{11}\text{B}$ records from the Monte Carlo approach, one of these calibrations was randomly chosen and applied to the whole time series (no calibration was chosen more than once). For each data point in the ensuing pH time series, the maximum probability of the distribution of the 10,000 pH estimates was determined with uncertainties given by the 2.5th, 16th, 84th and 97.5th percentiles.

To calculate $p_{\text{CO}_2}^{\text{sw}}$, another variable of the ocean carbonate system apart from pH is required²⁰. Here we use total alkalinity. Both pH and $p_{\text{CO}_2}^{\text{sw}}$ are governed by the ratio of dissolved inorganic carbon to total alkalinity in seawater, and so pH changes are proportional to changes in $p_{\text{CO}_2}^{\text{sw}}$ (refs 20, 82). Although our knowledge of total alkalinity is largely uncertain⁸³, modelling studies provide useful constraints on glacial–interglacial (GIG) total alkalinity change^{84,85}. The GIG total alkalinity change is $+120 \mu\text{mol kg}^{-1}$ in the ‘seven-box model’ of ref. 84, and is $+140 \mu\text{mol kg}^{-1}$ in ref. 85. There is no available information, however, on the secular evolution of total alkalinity during the deglaciation.

We calculated $p_{\text{CO}_2}^{\text{sw}}$ using the equations of ref. 20 and the ‘seacarb’ package⁸⁶ in R. The relevant input parameters were taken from the 10,000 Monte Carlo pH simulations above with the addition of total alkalinity, which randomly varied for each data point in the simulation from ‘modern minus $25 \mu\text{mol kg}^{-1}$ ’ to ‘modern plus $125 \mu\text{mol kg}^{-1}$ ’ with a ‘flat’ probability (that is, an equal probability of total alkalinity being any value between these extremes at any point of the record). This approach avoids ascribing weight to any particular total alkalinity value and fully explores the likely range given the available, model-based, constraints.

It is important to note that $p_{\text{CO}_2}^{\text{sw}}$ estimates are most determined by the reconstructed pH and that total alkalinity has little influence. For example, assuming that total alkalinity in both ODP1238 and PS2498-1 records is constant at either ‘modern plus $125 \mu\text{mol kg}^{-1}$ ’ or ‘modern minus $25 \mu\text{mol kg}^{-1}$ ’ modifies reconstructed $p_{\text{CO}_2}^{\text{sw}}$ only by a maximum of $\sim 22 \mu\text{atm}$ (Extended Data Fig. 9). Similarly, temperature and salinity have little effect on our pH and $p_{\text{CO}_2}^{\text{sw}}$ calculations, which are overwhelmingly determined by $\delta^{11}\text{B}$ and its associated uncertainties. The uncertainties associated with these parameters for PS2498-1 and ODP1238 are respectively at most ± 19 and $\pm 9 \mu\text{atm}$ for calibration uncertainty, ± 10 and $\pm 12 \mu\text{atm}$ for total alkalinity, ± 8 and $\pm 15 \mu\text{atm}$ for salinity, and ± 9 and $\pm 13 \mu\text{atm}$ for temperature (Extended Data Fig. 9).

To calculate ‘equilibrium pH’, that is, surface seawater pH expected at each core location if waters had remained in equilibrium with the contemporaneous atmosphere during the past 25 kyr (Fig. 2b, g), we used interpolated ice-core p_{CO_2} data^{5,87,88} and the calculated temperature, salinity and total alkalinity for each of our samples (see above). Calculations were made using the ‘seacarb’ package⁸⁶ in R with the constants of refs 67, 75, 76.

Here we define Δp_{CO_2} as the partial pressure of CO_2 in seawater minus the partial pressure of CO_2 in the atmosphere:

$$\Delta p_{\text{CO}_2} = p_{\text{CO}_2}^{\text{sw}} - p_{\text{CO}_2}^{\text{atm}}$$

Our reconstructed Δp_{CO_2} for the most recent samples in PS2498-1 and ODP1238 are -3 ± 16 and $+27 \pm 17 \mu\text{atm}$, respectively, which agree within uncertainties with modern mean annual Δp_{CO_2} from nearby locations (-15 ± 8 and $+45 \pm 8 \mu\text{atm}$, respectively; ref. 11). Because Δp_{CO_2} is a comparison between our p_{CO_2} records and the Antarctic ice-core p_{CO_2} record⁵, the resulting Δp_{CO_2} can be affected by the chronology used, especially in periods of rapid $p_{\text{CO}_2}^{\text{atm}}$ increase. To account for this, the uncertainty in Δp_{CO_2} due to our age models has been approximated by propagating a ± 0.5 kyr uncertainty in our records using a Monte Carlo approach. The resulting uncertainty in calculated Δp_{CO_2} is at most $\pm 12 \mu\text{atm}$ for PS2498-1 and $\pm 7 \mu\text{atm}$ for ODP1238 (Extended Data Fig. 9).

To derive CO_2 flux between the ocean and the atmosphere, Δp_{CO_2} and the sea–air gas transfer rate (normally parameterized as a function of wind speed¹¹) need to be considered. However, a visual comparison between our Fig. 1 and fig. 13 in ref. 11 illustrates that areas of high Δp_{CO_2} tend to correspond to areas of sea–air CO_2 flux. Consequently, and in agreement with previous literature, in our interpretations we assume that a positive Δp_{CO_2} implies CO_2 outgassing from the ocean to the atmosphere.

Smoothing of the records and Monte Carlo simulations. The $\delta^{11}\text{B}$ -derived and $\delta^{13}\text{C}$ records presented in this study have been smoothed by fitting a non-parametric regression (LOESS function). Smoothing was performed in R with the degree of smoothing (the ‘span’ term) optimized using both ‘general cross-validation’ and ‘leave-one-out cross-validation’ methods⁸⁹. These approaches identified nearly identical optimal degrees of smoothing for the $\delta^{11}\text{B}$ -derived records (span = 0.25 for PS2498-1 and 0.23 for ODP1238), and the most likely smoothed fit to the data was obtained using a probabilistic approach. To achieve this a LOESS function was fitted to each of the 10,000 realizations (for pH, $p_{\text{CO}_2}^{\text{sw}}$ and Δp_{CO_2}) generated using the Monte Carlo approach described above, and for each time step the distribution of smoothed lines was examined and the maximum probability and the 2.5th, 16th, 84th and 97.5th percentiles were determined⁹⁰. This approach fully accounts for the uncertainty in all of the input parameters, and provides an uncertainty in the most likely smoothed fit to the data. Uncertainty envelopes in the Δp_{CO_2} records also include the uncertainty associated with our age models.

Hydrographic control on seawater pH in the SAA. Millennial-scale reorganizations of the interhemispheric climate^{91,92} were accompanied by meridional shifts of the oceanic fronts of the Antarctic circumpolar current²⁷. PS2498-1 sits between the subtropical front and the sub-Antarctic front⁹³, the migrations of which may have influenced surface ocean hydrography and carbonate system parameters at this location. The main negative shifts in the PS2498-1 *G. bulloides* $\delta^{11}\text{B}$ record (and pH) generally occurred during warming episodes, as revealed by the co-registered *G. bulloides* Mg/Ca (Fig. 2d). This would agree with the contemporaneous southward shifts of the subtropical front at a nearby location²⁷ that are interpreted as the response of the South Atlantic to the main episodes of interhemispheric climate changes (for example during Heinrich Stadial 1 and the Younger Dryas). Today, surface ocean pH decreases southwards across the Antarctic circumpolar current⁹⁴, implying that a southward shift of the subtropical front would have plausibly caused a pH increase (as opposed to the observed decrease). Hence, the pH variability at PS2498-1 does not reflect reorganizations of the regional oceanic fronts, because these, at most, would have worked to counteract or damp, or both, the reconstructed acidification events.

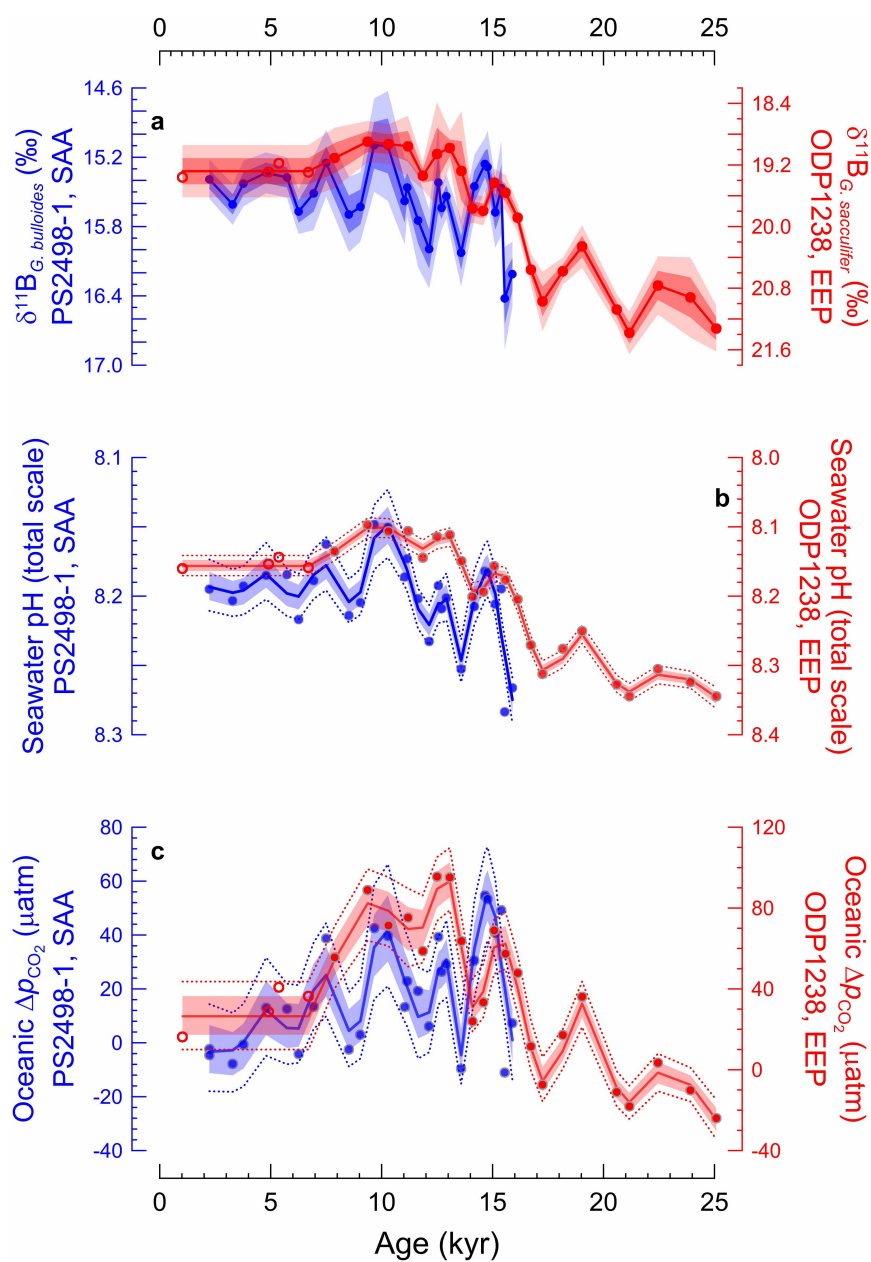
Intertropical convergence zone migrations in the EEP. The mean latitudinal position of the intertropical convergence zone (ITCZ) probably migrated southwards during Northern Hemisphere cold phases^{28,95–97}. ITCZ southward shifts during boreal winter currently coincide with decreased upwelling in the EEP (and vice versa). The timing of the main shifts towards low $\delta^{11}\text{B}$ and pH (high Δp_{CO_2}) in the ODP1238 matched Northern Hemisphere cooling events and attendant ITCZ southward displacements²⁸. However, they cannot be explained by variations in EEP upwelling rates because reduced upwelling would cause lower Δp_{CO_2} , as opposed to the observed higher Δp_{CO_2} . In addition, the appearance of low-pH waters in the ODP1238 deglacial record are associated with considerably increased temperatures, both in surface and, especially, in thermocline waters⁹⁸ (Fig. 2i and Extended Data Fig. 10), which cannot be fully explained by deglacial warming. In fact, these positive temperature excursions at ODP1238 exceeded Holocene temperatures.

We therefore conclude that the ITCZ shifts cannot explain the observed carbonate chemistry and temperature changes documented in our EEP records.

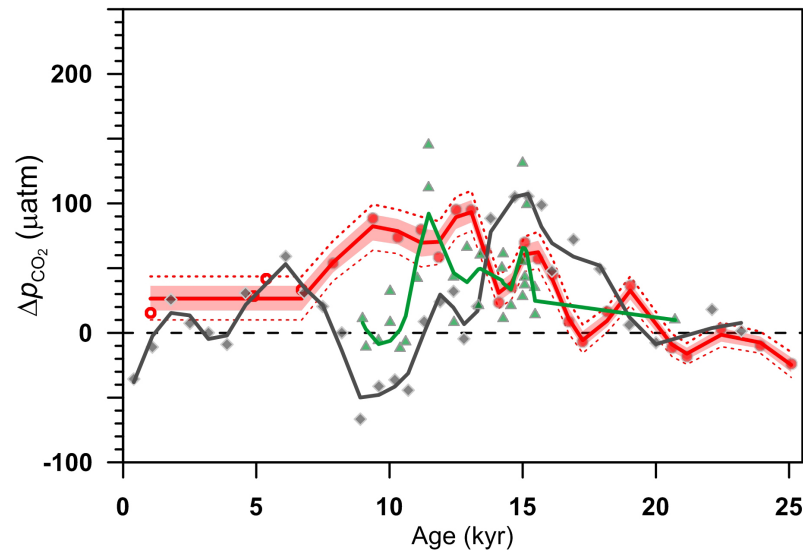
Oceanic tunnelling. Oceanic tunnelling is the transfer of intermediate waters (Antarctic intermediate waters or sub-Antarctic mode waters), and of the geochemical and thermal signals that they transport, from the Southern Ocean to the equatorial Pacific^{9,41}. These intermediate waters entrain upwelled deep waters during their formation in the high-latitude Southern Ocean, and during the last deglaciation their chemical characteristics would therefore have been influenced by changes in upwelling around Antarctica. After formation, they spread northwards (being modified along their path²⁴) and feed the Pacific equatorial undercurrent, an eastwards-flowing subsurface current that transports thermocline waters from the western Pacific along the equator^{99,100} and upwells in the EEP¹⁸. Therefore, this intermediate water route provides an efficient connection¹⁰¹ between the Southern Ocean and the EEP^{4,7,41}.

32. Gersonde, R. *et al.* Last glacial sea surface temperatures and sea-ice extent in the Southern Ocean (Atlantic-Indian sector): a multiproxy approach. *Paleoceanography* **18**, 1061 (2003).
33. Mackensen, A., Rudolph, M. & Kuhn, G. Late Pleistocene deep-water circulation in the subantarctic eastern Atlantic. *Global Planet. Change* **30**, 197–229 (2001).
34. Mix, A. *et al.* *Proceedings ODP, Initial Reports 202* (Ocean Drilling Program, 2003).
35. Stuiver, M. & Reimer, P. J. Extended ¹⁴C database and revised CALIB radiocarbon calibration program. *Radiocarbon* **35**, 215–230 (1993).
36. Stuiver, M., Reimer, P. J. & Reimer, R. W. CALIB Radiocarbon Calibration <http://calib.qub.ac.uk/calib/> (2005).
37. Key, R. M. *et al.* A global ocean carbon climatology: results from Global Data Analysis Project (GLODAP). *Glob. Biogeochem. Cycles* **18**, GB4031 (2004).
38. Siani, G. *et al.* Carbon isotope records reveal precise timing of enhanced Southern Ocean upwelling during the last deglaciation. *Nature Commun.* **4**, 2758 (2013).
39. Skinner, L. C., Fallon, S., Waelbroeck, C., Michel, E. & Barker, S. Ventilation of the deep Southern Ocean and deglacial CO₂ rise. *Science* **328**, 1147–1151 (2010).
40. Lea, D. W. *et al.* Paleoclimate history of Galápagos surface waters over the last 135,000 yr. *Quat. Sci. Rev.* **25**, 1152–1167 (2006).
41. Pena, L. D., Cacho, I., Ferretti, P. & Hall, M. A. El Niño–Southern Oscillation-like variability during glacial terminations and interlatitudinal teleconnections. *Paleoceanography* **23**, PA3101 (2008).
42. Marchitto, T. M., Lehman, S. J., Ortiz, J. D., Fluckiger, J. & van Geen, A. Marine radiocarbon evidence for the mechanism of deglacial atmospheric CO₂ rise. *Science* **316**, 1456–1459 (2007).
43. Stott, L., Southon, J., Timmermann, A. & Koutavas, A. Radiocarbon age anomaly at intermediate water depth in the Pacific Ocean during the last deglaciation. *Paleoceanography* **24**, PA2223 (2009).
44. Doss, W. & Marchitto, T. M. Global deep ocean sequestration of CO₂ driven by the eastern equatorial Pacific biologic pump. *Earth Planet. Sci. Lett.* **377–378**, 43–54 (2013).
45. Niebler, H. S. & Gersonde, R. A planktic foraminiferal transfer function for the southern South Atlantic Ocean. *Mar. Micropaleontol.* **34**, 213–234 (1998).
46. Mortyn, P. G. & Charles, C. D. Planktonic foraminiferal depth habitat and $\delta^{18}\text{O}$ calibrations: plankton tow results from the Atlantic sector of the Southern Ocean. *Paleoceanography* **18**, 1037 (2003).
47. Thunell, R. C., Curry, W. B. & Honjo, S. Seasonal variation in the flux of planktonic foraminifera: time series sediment trap results from the Panama Basin. *Earth Planet. Sci. Lett.* **64**, 44–55 (1983).
48. Thunell, R. C. & Reynolds, L. A. Sedimentation of planktonic foraminifera: seasonal changes in species flux in the Panama Basin. *Micropaleontology* **30**, 243–262 (1984).
49. Curry, W. B., Thunell, R. C. & Honjo, S. Seasonal-changes in the isotopic composition of planktonic foraminifera collected in Panama Basin sediment traps. *Earth Planet. Sci. Lett.* **64**, 33–43 (1983).
50. Fairbanks, R. G., Sverdrup, M., Free, R., Wiebe, P. H. & Be, A. W. H. Vertical distribution and isotopic fractionation of living planktonic foraminifera from the Panama Basin. *Nature* **298**, 841–844 (1982).
51. Faul, K. L., Ravelo, A. C. & Delaney, M. L. Reconstructions of upwelling, productivity, and photic zone depth in the eastern equatorial Pacific Ocean using planktonic foraminiferal stable isotopes and abundances. *J. Foraminiferal Res.* **30**, 110–125 (2000).
52. Bé, A. W. H. Gametogenic calcification in a spinose planktonic foraminifer, *Globigerinoides sacculifer* (Brady). *Mar. Micropaleontol.* **5**, 283–310 (1980).
53. Spero, H. J., Mielke, K. M., Kalve, E. M., Lea, D. W. & Pak, D. K. Multispecies approach to reconstructing eastern equatorial Pacific thermocline hydrography during the past 360 kyr. *Paleoceanography* **18**, 1022 (2003).
54. Caron, D. A., Anderson, O. R., Lindsey, J. L., Faber, W. W. & Lim, E. L. Effects of gametogenesis on test structure and dissolution of some spinose planktonic-foraminifera and implications for test preservation. *Mar. Micropaleontol.* **16**, 93–116 (1990).
55. Niebler, H. S. *Stable Isotopes Measured on Globigerina Bulloides of Sediment Core PS2498-1* <http://doi.pangaea.de/10.1594/PANGAEA.55892> (2004).
56. Barker, S., Greaves, M. & Elderfield, H. A study of cleaning procedures used for foraminiferal Mg/Ca paleothermometry. *Geochem. Geophys. Geosyst.* **4**, 8407 (2003).
57. Yu, J. M., Elderfield, H., Greaves, M. & Day, J. Preferential dissolution of benthic foraminiferal calcite during laboratory reductive cleaning. *Geochem. Geophys. Geosyst.* **8**, Q06016 (2007).
58. Rae, J. W. B., Foster, G. L., Schmidt, D. N. & Elliott, T. Boron isotopes and B/Ca in benthic foraminifera: proxies for the deep ocean carbonate system. *Earth Planet. Sci. Lett.* **302**, 403–413 (2011).
59. Anand, P., Elderfield, H. & Conte, M. H. Calibration of Mg/Ca thermometry in planktonic foraminifera from a sediment trap time series. *Paleoceanography* **18**, 1050 (2003).
60. Mashiotto, T. A., Lea, D. W. & Spero, H. J. Glacial-interglacial changes in Subantarctic sea surface temperature and $\delta^{18}\text{O}$ -water using foraminiferal Mg. *Earth Planet. Sci. Lett.* **170**, 417–432 (1999).
61. Boyle, E. A. & Keigwin, L. D. Comparison of Atlantic and Pacific paleochemical records for the last 215,000 years: changes in deep ocean circulation and chemical inventories. *Earth Planet. Sci. Lett.* **76**, 135–150 (1985).
62. Hennehan, M. J. *et al.* Calibration of the boron isotope proxy in the planktonic foraminifera *Globigerinoides ruber* for use in palaeo-CO₂ reconstruction. *Earth Planet. Sci. Lett.* **364**, 111–122 (2013).
63. Catanzaro, E. J. *et al.* *Boric Acid: Isotopic and Assay Standard Reference Material* (National Bureau of Standards Spec. Publ. **260**, US Govt Printing Office, 1970).
64. Klochko, K., Kaufman, A. J., Yao, W., Byrne, R. H. & Tossell, J. A. Experimental measurement of boron isotope fractionation in seawater. *Earth Planet. Sci. Lett.* **248**, 276–285 (2006).
65. Hemming, N. G. & Hanson, G. N. Boron isotopic composition and concentration in modern marine carbonates. *Geochim. Cosmochim. Acta* **56**, 537–543 (1992).
66. Hemming, N. G., Reeder, R. J. & Hanson, G. N. Mineral-fluid partitioning and isotopic fractionation of boron in synthetic calcium carbonate. *Geochim. Cosmochim. Acta* **59**, 371–379 (1995).
67. Dickson, A. G. Thermodynamics of the dissociation of boric-acid in synthetic seawater from 273.15 to 318.15 K. *Deep-Sea Res.* **37**, 755–766 (1990).
68. Foster, G. L., Pogge von Strandmann, P. A. E. & Rae, J. W. B. Boron and magnesium isotopic composition of seawater. *Geochem. Geophys. Geosyst.* **11**, Q08015 (2010).
69. Hennehan, M. J. *Ground-Truthing the Boron-Based Proxies*. PhD thesis, Univ. Southampton (2013).
70. Prebble, J. G. *et al.* An expanded modern dinoflagellate cyst dataset for the Southwest Pacific and Southern Hemisphere with environmental associations. *Mar. Micropaleontol.* **101**, 33–48 (2013).
71. Key, R. M. *et al.* The CARINA data synthesis project: introduction and overview. *Earth Syst. Sci. Data* **2**, 105–121 (2010).
72. Lee, K. *et al.* Global relationships of total alkalinity with salinity and temperature in surface waters of the world's oceans. *Geophys. Res. Lett.* **33**, L19605 (2006).
73. Gloor, M. *et al.* A first estimate of present and preindustrial air-sea CO₂ flux patterns based on ocean interior carbon measurements and models. *Geophys. Res. Lett.* **30**, 1010 (2003).
74. van Heuven, S., Pierrot, D., Rae, J. W. B., Lewis, E. & Wallace, D. W. R. *MATLAB Program Developed for CO₂ System Calculations* http://cdiac.ornl.gov/ftp/co2sys/CO2SYS_calc_MATLAB_v1.1/ (Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, 2011).
75. Lueker, T. J., Dickson, A. G. & Keeling, C. D. Ocean pCO₂ calculated from dissolved inorganic carbon, alkalinity, and equations for K₁ and K₂: validation based on laboratory measurements of CO₂ in gas and seawater at equilibrium. *Mar. Chem.* **70**, 105–119 (2000).
76. Lee, K. *et al.* The universal ratio of boron to chlorinity for the North Pacific and North Atlantic oceans. *Geochim. Cosmochim. Acta* **74**, 1801–1811 (2010).
77. Foster, G. L., Lear, C. H. & Rae, J. W. B. The evolution of pCO₂, ice volume and climate during the middle Miocene. *Earth Planet. Sci. Lett.* **341–344**, 243–254 (2012).
78. Sanyal, A., Bijma, J., Spero, H. & Lea, D. W. Empirical relationship between pH and the boron isotopic composition of *Globigerinoides sacculifer*: implications for the boron isotope paleo-pH proxy. *Paleoceanography* **16**, 515–519 (2001).
79. Foster, G. L. *et al.* Interlaboratory comparison of boron isotope analyses of boric acid, seawater and marine CaCO₃ by MC-ICPMS and NTIMS. *Chem. Geol.* **358**, 1–14 (2013).
80. Sanyal, A., Hemming, N. G., Hanson, G. N. & Broecker, W. S. Evidence for a higher pH in the glacial ocean from boron isotopes in foraminifera. *Nature* **373**, 234–236 (1995).
81. Schmidt, G. A. Error analysis of paleosalinity calculations. *Paleoceanography* **14**, 422–429 (1999).
82. Hain, M. P., Sigman, D. M. & Haug, G. H. in *Treatise on Geochemistry* (eds Holland, H. D. & Turekian, K. K.) 485–517 (Elsevier, 2014).
83. Yu, J., Foster, G. L., Elderfield, H., Broecker, W. S. & Clark, E. An evaluation of benthic foraminiferal B/Ca and $\delta^{11}\text{B}$ for deep ocean carbonate ion and pH reconstructions. *Earth Planet. Sci. Lett.* **293**, 114–120 (2010).
84. Toggweiler, J. R. Variation of atmospheric CO₂ by ventilation of the ocean's deepest water. *Paleoceanography* **14**, 571–588 (1999).
85. Hain, M. P., Sigman, D. M. & Haug, G. H. Carbon dioxide effects of Antarctic stratification, North Atlantic Intermediate Water formation, and subantarctic nutrient drawdown during the last ice age: diagnosis and synthesis in a geochemical box model. *Glob. Biogeochem. Cycles* **24**, GB4023 (2010).
86. R Development Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2013).
87. Monnin, E. *et al.* Atmospheric CO₂ concentrations over the last glacial termination. *Science* **291**, 112–114 (2001).
88. Laurantou, A. *et al.* Constraint of the CO₂ rise by new atmospheric carbon isotopic measurements during the last deglaciation. *Glob. Biogeochem. Cycles* **24**, GB2015 (2010).

89. Chandler, R. & Scott, M. *Statistical Methods for Trend Detection and Analysis in the Environmental Sciences* (Wiley, 2011).
90. Rohling, E. J. *et al.* Sea-level and deep-sea-temperature variability over the past 5.3 million years. *Nature* **508**, 477–482 (2014).
91. Stocker, T. F. The seesaw effect. *Science* **282**, 61–62 (1998).
92. Barbante, C. *et al.* One-to-one coupling of glacial climate variability in Greenland and Antarctica. *Nature* **444**, 195–198 (2006).
93. Orsi, A. H., Whitworth, T. & Nowlin, W. D. On the meridional extent and fronts of the Antarctic Circumpolar Current. *Deep Sea Res. I* **42**, 641–673 (1995).
94. González-Dávila, M. *et al.* Carbonate system in the water masses of the Southeast Atlantic sector of the Southern Ocean during February and March 2008. *Biogeosciences* **8**, 1401–1413 (2011).
95. Arbuszewski, J. A., Demenocal, P. B., Cleroux, C., Bradtmiller, L. & Mix, A. Meridional shifts of the Atlantic intertropical convergence zone since the Last Glacial Maximum. *Nature Geosci.* **6**, 959–962 (2013).
96. Koutavas, A. & Lynch-Stieglitz, J. in *Hadley Circulation: Present, Past and Future* Vol. 21 (eds Diaz, H. F. & Bradley, R. S.) 347–369 (Kluwer, 2004).
97. Dubois, N., Kienast, M., Normandeau, C. & Herbert, T. D. Eastern equatorial Pacific cold tongue during the Last Glacial Maximum as seen from alkenone paleothermometry. *Paleoceanography* **24**, PA4207 (2009).
98. Sadekov, A. Y. *et al.* Palaeoclimate reconstructions reveal a strong link between El Niño–Southern Oscillation and Tropical Pacific mean state. *Nature Commun.* **4**, 2692 (2013).
99. Lukas, R. The termination of the equatorial undercurrent in the eastern Pacific. *Prog. Oceanogr.* **16**, 63–90 (1986).
100. Tsuchiya, M., Lukas, R., Fine, R. A., Firing, E. & Lindstrom, E. Source waters of the Pacific equatorial undercurrent. *Prog. Oceanogr.* **23**, 101–147 (1989).
101. Fine, R. A., Lukas, R., Bingham, F. M., Warner, M. J. & Gammon, R. H. The western equatorial Pacific: a water mass crossroads. *J. Geophys. Res.* **99**, 25063–25080 (1994).

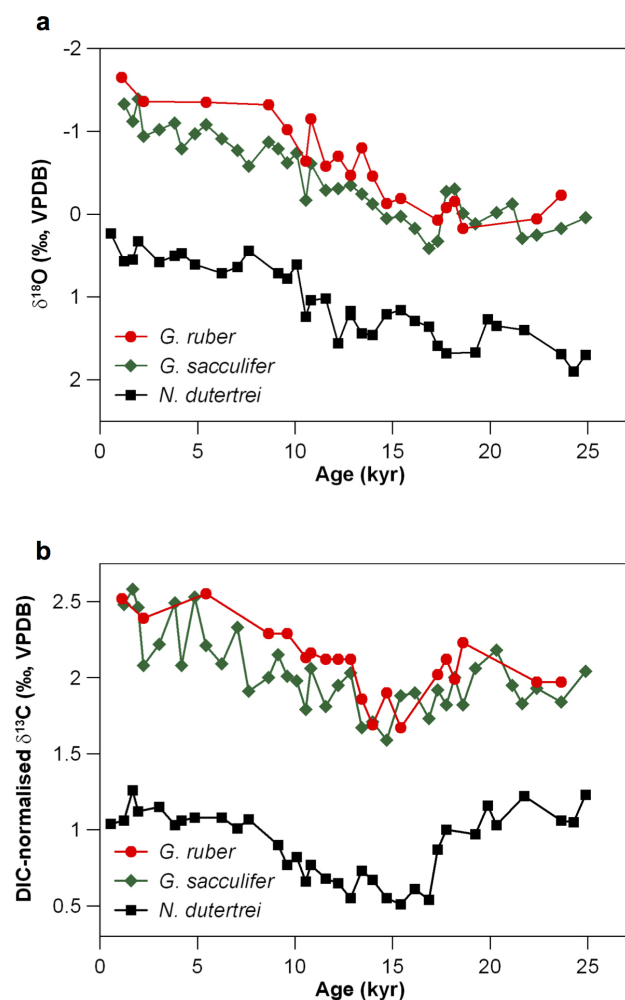


Extended Data Figure 1 | Comparison of PS2498-1 (blue) and ODP1238 (red) records. a, $\delta^{11}\text{B}$. b, pH. c, Δp_{CO_2} .



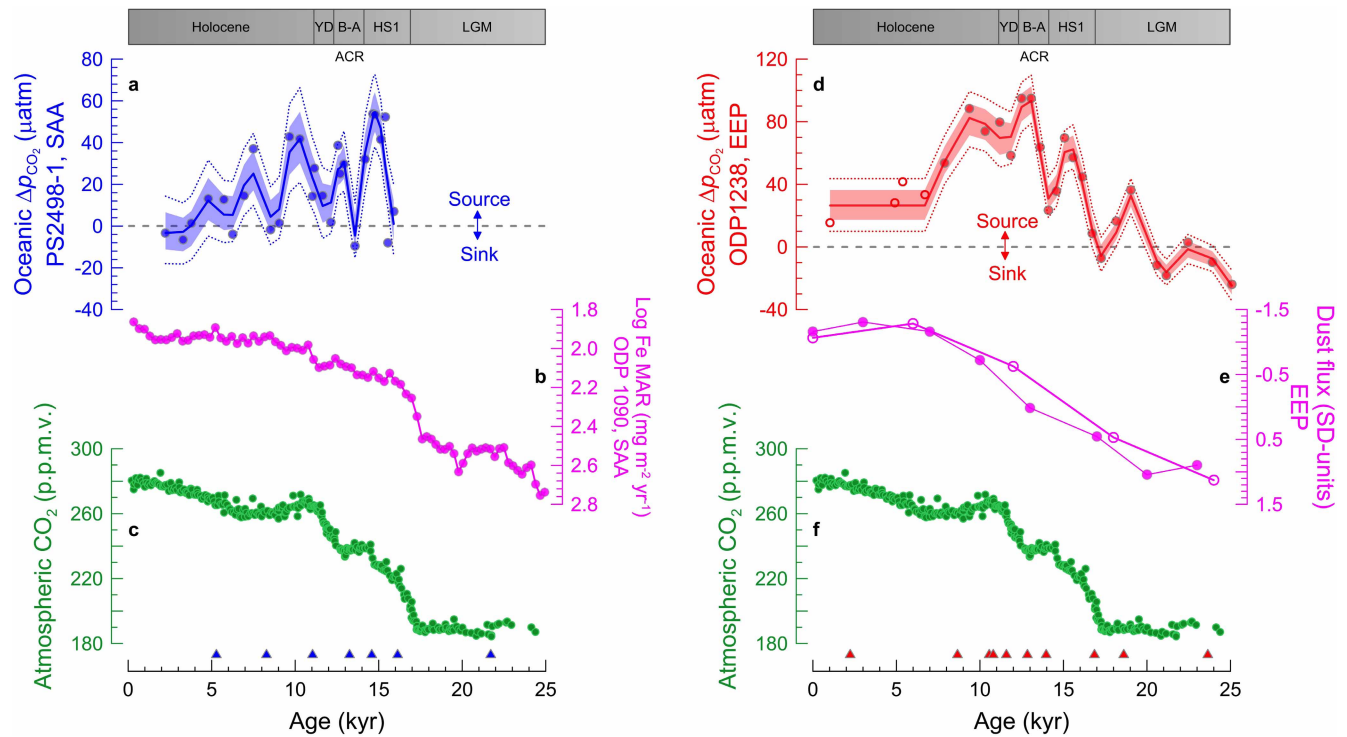
Extended Data Figure 2 | $\delta^{11}\text{B}$ -derived Δp_{CO_2} compilation for the equatorial Pacific during the last deglaciation and Holocene. Foraminifera-based record from the western equatorial Pacific²¹ (grey), *Porites* coral-based record from the central equatorial Pacific^{22,23} (as published in ref. 22, green), and foraminifera-based record from the EEP (this study, red). The records of refs 21, 22 have been smoothed by fitting a LOESS function with degrees of

smoothing (span) of 0.2 and 0.4, respectively (Methods), to allow a better comparison with the ODP1238 record (see main text). ODP1238 is located in the EEP, and therefore represents a direct record of upwelling of CO_2 -rich waters, while the signal at central and western equatorial sites may have been modified during the westward transit of waters by, for example, equilibration with the atmosphere and/or nutrient utilization by the biological pump.



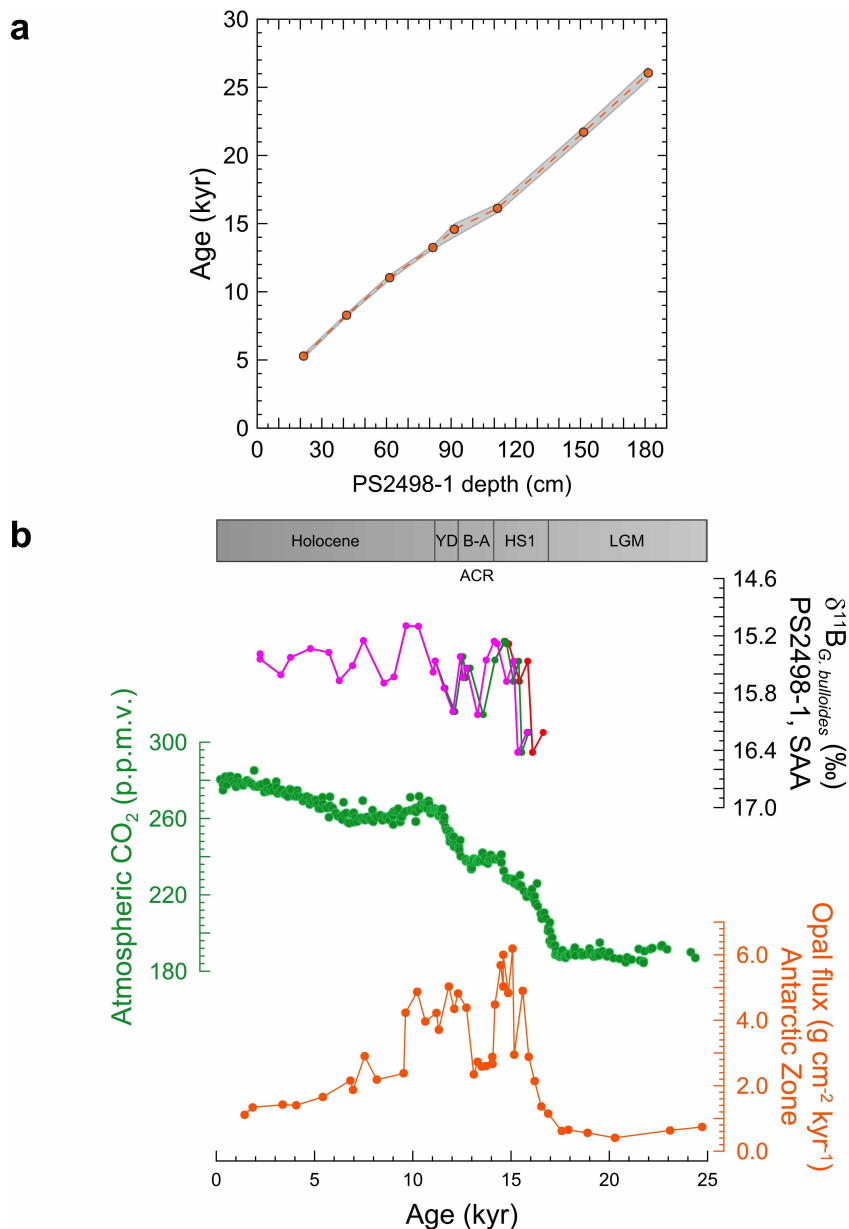
Extended Data Figure 3 | Planktic $\delta^{18}\text{O}$ and $\delta^{13}\text{C}$ records from ODP1238.

a, Planktic $\delta^{18}\text{O}$ records from ODP1238. Red, *G. ruber* sensu stricto (ss) 250–355 μm ; green, *G. sacculifer* (mixed morphotypes) 355–425 μm ; black, *N. dutertrei* 355–500 μm . **b**, Planktic $\delta^{13}\text{C}$ records from ODP1238. To facilitate comparison between species, $\delta^{13}\text{C}$ data has been normalized⁵³. Red, *G. ruber* ss 250–355 μm ; green, *G. sacculifer* (mixed morphotypes) 355–425 μm ; black, *N. dutertrei* 355–500 μm .



Extended Data Figure 4 | Δp_{CO_2} records from the SAA and the EEP during the last deglaciation, compared with indicators of dust input. **a–c**, SAA; **d–f**, EEP. **a**, $\delta^{11}\text{B}$ -derived Δp_{CO_2} in core PS2498-1. **b**, Logarithm of the mass accumulation rates (MAR) of iron (Fe) in the sub-Antarctic site ODP1090¹³.

c, f, $p_{\text{CO}_2}^{\text{atm}}$ measured on a suite of Antarctic ice cores⁵. **d**, $\delta^{11}\text{B}$ -derived Δp_{CO_2} in core ODP1238. **e**, Dust fluxes in the EEP²⁹. Dust fluxes from ref. 29 were de-meaned and divided by their own standard deviation, and are displayed in standard deviation units.



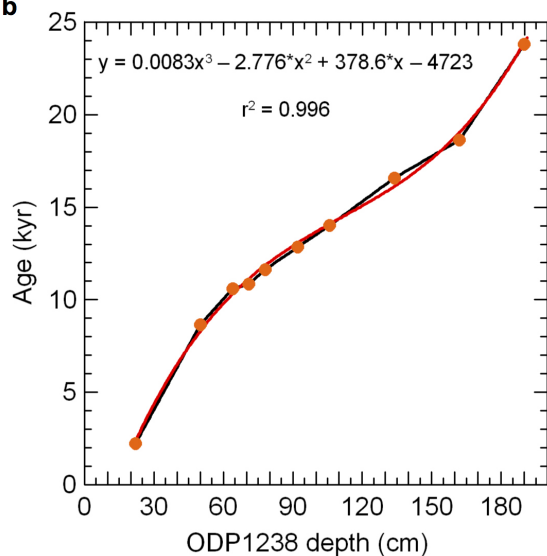
Extended Data Figure 5 | Age model for PS2498-1. **a**, Chronology of SAA core PS2498-1. Carbon-14 calendar age/depth relationships in core PS2498-1. Grey shading indicates 95% confidence limits of calendar ages. **b**, PS2498-1 *G. bulloides* $\delta^{11}\text{B}$ record plotted using the different chronologies described in Methods and compared with atmospheric CO_2 (green; refs 5, 87, 88) and with

Antarctic opal flux⁴ (orange) records. Red, constant $\Delta R = 300$ yr; green, $\Delta R = 900$ yr for intervals older than 16 kyr and $\Delta R = 300$ yr for younger intervals⁶; magenta, variable ΔR correction³⁸ (ranging between 500 and 900 yr between 13 and 16 kyr ago).

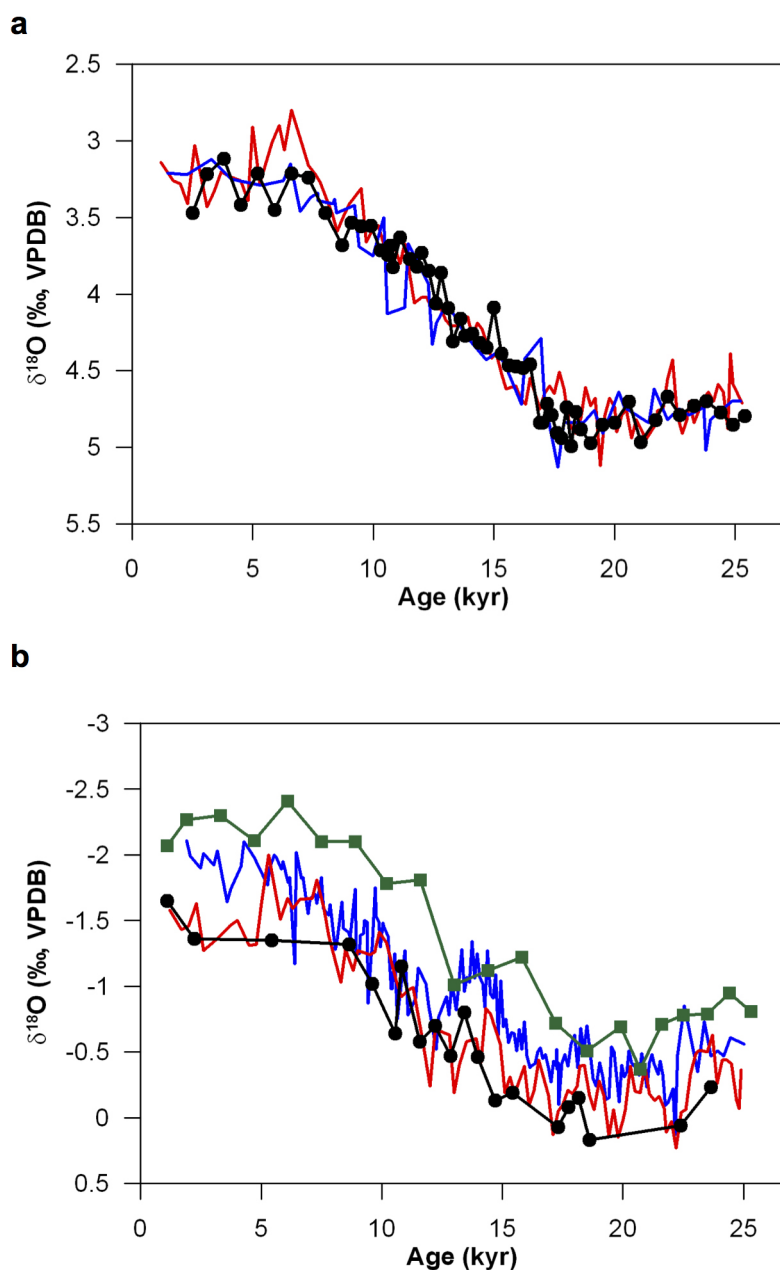
a

Core-depth (cm)	¹⁴ C age (yr)	±1σ	Calendar age (yr)	±1σ
21-23	2625	25	2230	60
49-51	8235	30	8650	75
63-65	9765	30	10590	60
70-72	9955	35	10830	100
77-79	10545	35	11620	160
91-93	11460	30	12840	70
105-107	12600	40	14010	80
133-135	14210	35	16580	130
161-163	15825	45	18630	70
189-191	20250	70	23800	130

b



Extended Data Figure 6 | Age model for ODP1238. **a**, Radiocarbon ages for ODP1238 determined from *N. dutertrei* tests at LLNL-CAMS. **b**, Chronology of EEP core ODP1238. Orange circles, calendar ages; black line, linear fit; red line, third-order polynomial fit (Methods).



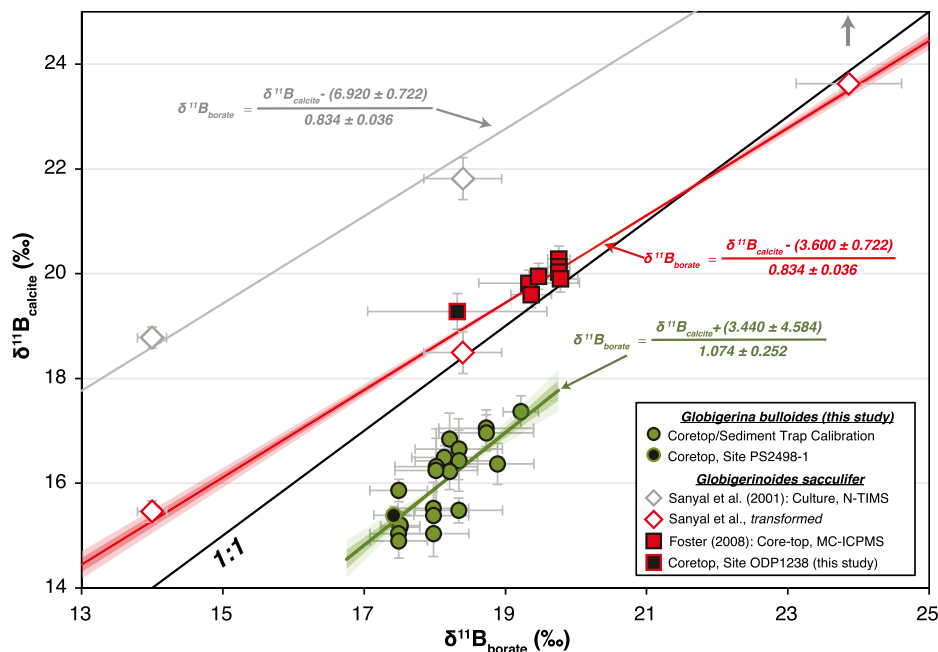
Extended Data Figure 7 | Benthic and planktic $\delta^{18}\text{O}$ stratigraphy for ODP1238. **a**, Benthic $\delta^{18}\text{O}$ stratigraphy for ODP1238 compared with other benthic $\delta^{18}\text{O}$ stratigraphies from EEP cores. Black circles, unpublished benthic $\delta^{18}\text{O}$ data for ODP1238 generated by J. F. McManus (LDEO, Columbia University); red line, site TR163-22⁴⁰; blue line, sites RC13-140, RC23-22 and

RC23-15⁴⁴. **b**, *Globigerinoides ruber* $\delta^{18}\text{O}$ stratigraphy for ODP1238 compared with other *G. ruber* stratigraphies from EEP cores. Black circles, ODP1238 (Methods); green squares, site TR163-19⁷; red line, site TR163-22⁴⁰; blue line, site ODP1240⁴¹.

a

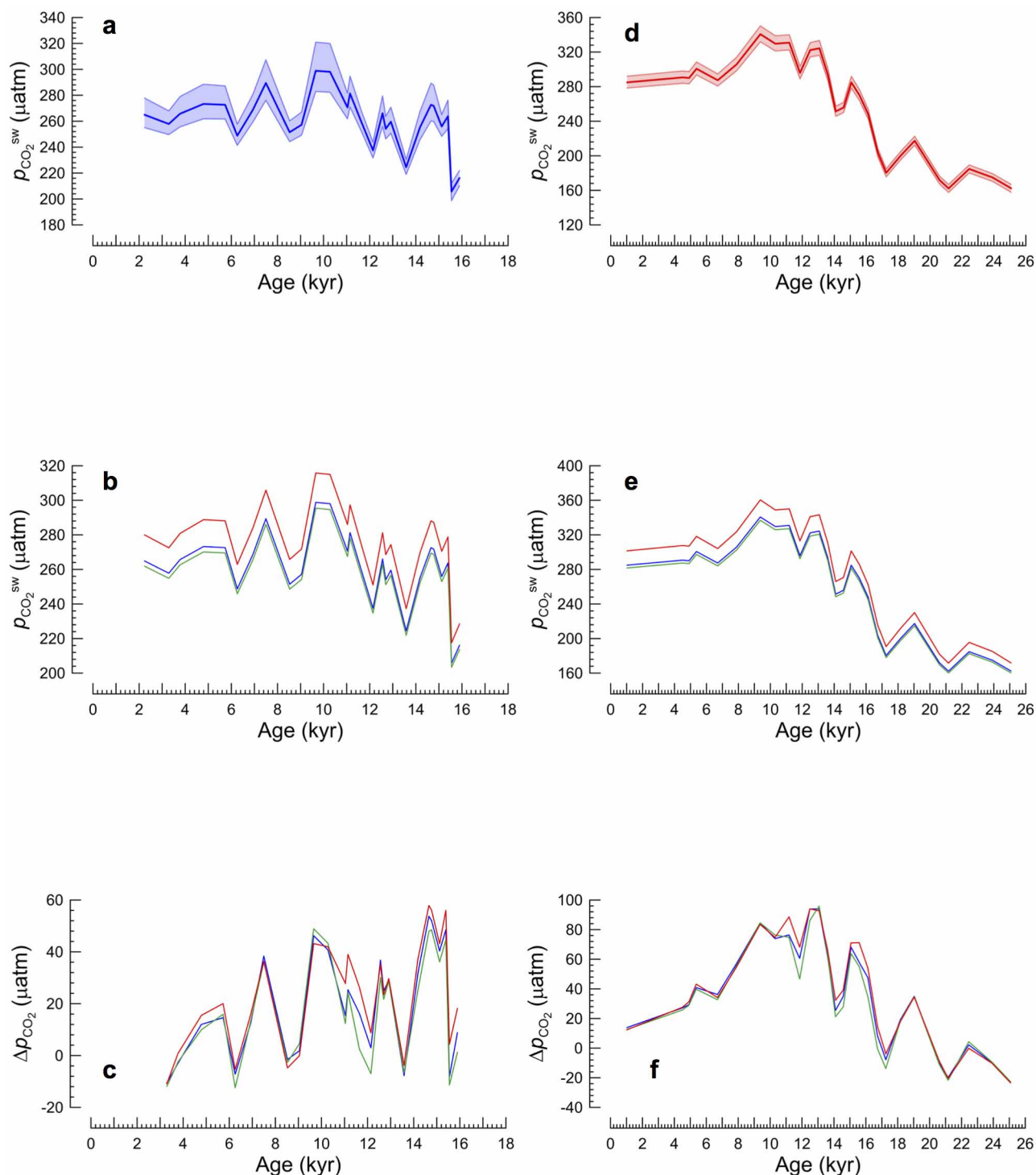
Site	Sample type	Size fraction (μm)	Latitude (°N)	Longitude (°E)	T (°C)	Salinity	pH	±2σ	pK _a	δ ¹¹ B _{calcite} (‰)	±2σ	δ ¹¹ B _{borate} (‰)	±2σ
F111	Core-top	355-400	-48.95	174.98	9.08	34.33	8.191	0.003	8.794	15.86	0.21	17.49	0.41
MC577-17B	Core-top	355-400	45.57	-17.40	15.32	35.70	8.188	0.002	8.709	16.65	0.58	18.34	0.62
F111	Core-top	250-300	-48.95	174.98	9.08	34.33	8.191	0.003	8.794	14.90	0.33	17.49	0.41
CAR22(Z)6	Sediment Trap	250-300	10.50	-64.66	24.17	36.70	8.066	0.018	8.597	16.23	0.35	18.21	0.11
MC577-17B	Core-top	250-300	45.57	-17.40	15.32	35.70	8.188	0.002	8.709	16.43	0.58	18.34	0.62
MC436	Core-top	300-355	39.80	-21.06	18.40	36.06	8.201	0.017	8.670	16.37	0.40	18.89	0.51
MC655	Core-top	300-355	38.42	5.40	23.23	37.38	8.161	0.010	8.605	17.36	0.31	19.22	0.25
F111	Core-top	300-355	-48.95	174.98	9.08	34.33	8.191	0.003	8.794	15.04	0.21	17.49	0.41
TAN1106/38	Core-top ('flattened' morphotype)	300-355	-49.69	165.07	9.78	34.49	8.186	0.003	8.783	15.16	0.26	17.51	0.28
TAN1106/38	Core-top ('kummerform' morphotype)	300-355	-49.69	165.07	9.78	34.49	8.186	0.003	8.783	15.19	0.27	17.51	0.28
CAR22(Z)6	Sediment Trap	300-355	10.50	-64.66	24.17	36.70	8.066	0.018	8.597	16.84	0.50	18.21	0.11
ODP1172C	Core-top	300-355	-43.96	149.93	13.72	35.02	8.196	0.004	8.732	16.49	0.23	18.14	0.47
MC577-17B	Core-top	300-355	45.57	-17.40	15.32	35.70	8.188	0.002	8.709	15.48	0.24	18.34	0.62
IODP1313	Core-top	300-355	41.00	-32.96	18.50	36.03	8.182	0.008	8.668	16.96	0.34	18.73	0.67
IODP1308	Core-top	300-355	49.88	-24.24	13.22	35.41	8.183	0.001	8.736	15.52	0.50	17.98	0.50
ODP 980	Core-top	300-355	55.49	-14.70	11.71	35.36	8.206	0.016	8.755	16.32	0.72	18.02	0.59
ODP 980	Core-top	250-300	55.49	-14.70	11.71	35.36	8.206	0.016	8.755	16.24	0.62	18.02	0.59
IODP 1313	Core-top	250-300	41.00	-32.96	18.50	36.03	8.182	0.008	8.668	17.05	0.35	18.73	0.67
IODP 1308	Core-top	250-300	49.88	-24.24	13.22	35.41	8.183	0.001	8.736	15.38	0.45	17.98	0.50
IODP 1308	Core-top	355-400	49.88	-24.24	13.22	35.41	8.183	0.001	8.736	15.04	0.44	17.98	0.50

b



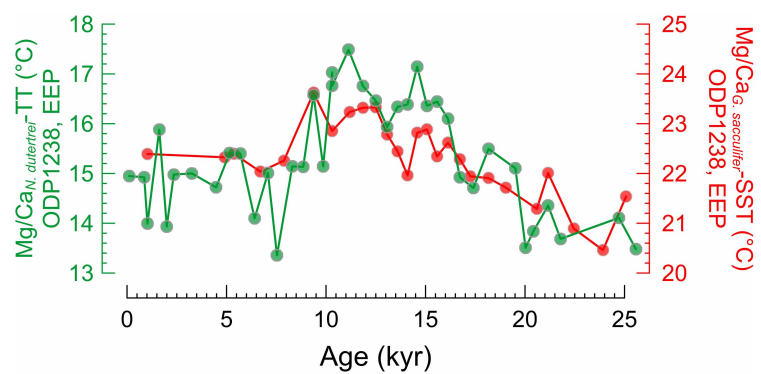
Extended Data Figure 8 | $\delta^{11}\text{B}$ -pH calibrations for *G. bulloides* and *G. sacculifer*. **a**, Data tabulated; **b**, Data plotted. The green symbols and text show a new calibration for *G. bulloides* (with associated 2σ uncertainties). Horizontal error bars for core-top samples are 2σ of intra-annual variability in calculated monthly $\delta^{11}\text{B}_{\text{borate}}$ and for sediment trap samples reflect the range of $\delta^{11}\text{B}_{\text{borate}}$ between December 2006 and February 2007. Vertical error bars represent the analytical reproducibility (2σ) as calculated using equation (1). The most recent PS2498-1 sample (2.2 kyr old) (black-filled circle) was not used in the calibration process, and is included to show its agreement with the calibration line. The red symbols and text show a calibration for *G. sacculifer* (with associated 2σ uncertainties). The calibration line incorporates both culture⁷⁸ (empty symbols) and core-top (red-filled symbols) data¹⁷. Culture data analysed by N-TIMS (grey symbols and text)⁷⁸ has been corrected by applying a laboratory offset of -3.32‰ (Methods) (the vertical grey arrow

indicates an original N-TIMS calibration data point that falls outside the plot area). The ODP1238 late-Holocene average (black-filled square) was not used to produce the calibration equation, and is included to show its agreement with the calibration line. Horizontal error bars for core-top samples are 2σ of intra-annual variability in calculated monthly $\delta^{11}\text{B}_{\text{borate}}$ and for culture samples represent quoted uncertainties⁷⁸ in pH. Vertical error bars represent quoted uncertainties in $\delta^{11}\text{B}$ measurements^{17,78} (2σ). To calculate monthly pH variations at ODP1238, the method described in the *G. bulloides* calibration section has been used⁶² (with total alkalinity derived using the total alkalinity/salinity/temperature relationship for the 'Equatorial upwelling Pacific Zone' in ref. 72). The black line denotes a 1:1 relationship, that is, a pH sensitivity equal to that of borate ion. Heavily and lightly shaded regions around calibration lines represent 1σ and 2σ uncertainties, respectively.



Extended Data Figure 9 | Effect of $\delta^{11}\text{B}$ -pH calibration, total alkalinity and chronological uncertainties in p_{CO_2} and Δp_{CO_2} records. **a**, PS2498-1 $\delta^{11}\text{B}$ -based p_{CO_2} record calculated with the *G. bulloides* calibration equation (thick blue line), and its associated 2σ uncertainty (blue shaded envelope). **b**, PS2498-1 $\delta^{11}\text{B}$ -based p_{CO_2} record assuming a constant total alkalinity of (i) modern values at PS2498-1 (blue), (ii) modern values minus $25 \mu\text{mol kg}^{-1}$ (green) and (iii) modern values plus $125 \mu\text{mol kg}^{-1}$ (red). **c**, PS2498-1 Δp_{CO_2} record calculated using (i) age derived from our age model (blue), (ii) age plus 0.5 kyr (green) and (iii) age minus 0.5 kyr (red). **d**, ODP1238 $\delta^{11}\text{B}$ -based

p_{CO_2} record calculated with the *G. sacculifer* calibration equation (thick red line), and its associated 2σ uncertainty (shaded red envelope). **e**, ODP1238 $\delta^{11}\text{B}$ -based p_{CO_2} record assuming a constant total alkalinity of (i) modern values at ODP1238 (blue), (ii) modern values minus $25 \mu\text{mol kg}^{-1}$ (green) and (iii) modern values plus $125 \mu\text{mol kg}^{-1}$ (red). **f**, ODP1238 Δp_{CO_2} record calculated using (i) age derived from our age model (blue), (ii) age plus 0.5 kyr (green) and (iii) age minus 0.5 kyr (red). Note the different horizontal and vertical axes in each panel.



Extended Data Figure 10 | *Globigerinoides sacculifer* Mg/Ca-based sea surface SST (red) and *N. dutertrei* Mg/Ca-based thermocline temperature (TT; green) at ODP1238.

Recharge of a subglacial lake by surface meltwater in northeast Greenland

Michael J. Willis^{1,2}, Bradley G. Herried³, Michael G. Bevis⁴ & Robin E. Bell⁵

In a warming climate, surface meltwater production on large ice sheets is expected to increase. If this water is delivered to the ice sheet base it may have important consequences for ice dynamics. For example, basal water distributed in a diffuse network can decrease basal friction^{1,2} and accelerate ice flow^{3–8}, whereas channelized basal water can move quickly to the ice margin, where it can alter fjord circulation and submarine melt rates^{9,10}. Less certain is whether surface meltwater can be trapped and stored in subglacial lakes beneath large ice sheets. Here we show that a subglacial lake in Greenland drained quickly, as seen in the collapse of the ice surface, and then refilled from surface meltwater input. We use digital elevation models from stereo satellite imagery and airborne measurements to resolve elevation changes during the evolution of the surface and basal hydrologic systems at the Flade Isblink ice cap in northeast Greenland. During the autumn of 2011, a collapse basin about 70 metres deep and about 0.4 cubic kilometres in volume formed near the southern summit of the ice cap as a subglacial lake drained into a nearby fjord. Over the next two years, rapid uplift of the floor of the basin (which is approximately 8.4 square kilometres in area) occurred as surface meltwater flowed into crevasses around the basin margin and refilled the subglacial lake. Our observations show that surface meltwater can be trapped and stored at the bed of an ice sheet. Sensible and latent heat released by this trapped meltwater could soften nearby colder basal ice¹¹ and alter downstream ice dynamics^{12,13}. Heat transport associated with meltwater trapped in subglacial lakes should be considered when predicting how ice sheet behaviour will change in a warming climate.

Studies of cryo-hydrologic systems in west Greenland suggest that supraglacial water that makes its way to the base of the ice sheet is flushed to the ocean through rapidly evolving subglacial drainage networks^{5–8}. The possibility of this water being captured¹⁴ and stored at the base has largely been neglected. Until recently, subglacial lakes remained undetected beneath Greenland. Two such lakes have now been identified beneath the northwesternmost part of the ice sheet¹⁵. These lakes are surprising because subglacial lake water is usually produced at the base of an ice sheet by geothermal and frictional melting^{16,17} but the ice above the lakes is too slow, too thin and too cold to support basal melting¹⁵. The water in these subglacial lakes is hypothesized to originate from nearby supraglacial lakes that periodically drain to the bed¹⁵. If Greenlandic subglacial lakes are replenished from the surface, the thermal environment around such lakes could be warmer than expected. Meltwater arriving at or near the base of the ice sheet will refreeze and release its latent heat¹¹. This process will warm the basal ice, which could change its rheology and thus affect downstream ice flow rates.

The ~8,500 km² Flade Isblink ice cap is situated close to Station Nord on the Princess Dagar peninsula (81.3° N, 15.0° W), northeast Greenland (Fig. 1). Numerical weather model output¹⁸ indicates the average annual surface temperature at the southern ice divide (the line at which ice flows either one way or another; see Extended Data Fig. 3) is about

–22 °C and the long-term accumulation rate is ~0.25 m of water equivalent per year. Using an estimated geothermal heat flux¹⁹ of ~60 mW m^{–2}, we calculate the temperature at the bed—about 540 m (Extended Data Fig. 1) beneath the ice divide—to be approximately –9 °C (ref. 20), far below the pressure melting temperature (about –0.5 °C). The base of the Flade Isblink ice cap is too cold to allow local production of basal meltwater. Surface meltwater is plentiful during the summers on the south-facing slopes of the southern part of the ice cap (see Fig. 1b). In 2006 supraglacial streams formed at high elevations just north of the ice divide. In the summers of 2006 to 2011 these northerly flowing meltwater streams always disappeared into a recurring moulin (81° 09′ 23″ N, 16° 36′ 04″ W).

In the late summer of 2011, a deep basin surrounded by crevasses formed on the surface of the Flade Isblink ice cap (Fig. 1, Extended Data Figs 2 and 3) just to the north of the ice divide. The basin formed over a 21-day interval during which clouds obscured 250-m resolution MODIS optical imagery. On 16 August 2011 the Flade Isblink ice cap's ice divide area had a smooth ice surface. On the next clear image on 6 September 2011, a 'mitten'-shaped collapse basin had formed. The 'mitten' is composed of two sub-basins separated by a shallow saddle. The larger, ~1.6-km-wide eastern main basin is centred on the location of the recurring moulin. A 2-m-high ice ridge aligned along the basin axis probably formed by compression during the initial slumping of the basin. We interpret the formation of the collapse basin to be the result of the sudden outburst of a subglacial lake that was overfilled by meltwater draining from the surface.

Along-track stereo satellite imagery collected 8 months later on 3 May 2012 provides the first topographic measurements of the collapse basin. At that time the basin had a maximum relief of more than 75 m (Figs 1 and 2) and an area of about 8.4 km². We set the pre-collapse surface elevations from a modified ICESat (Ice, Cloud and land Elevation Satellite)-controlled InSAR (INterferometric Synthetic Aperture Radar) digital elevation model (DEM)²¹ and calculate the basin volume to be ~0.4 km³. This basin volume is similar to many of the active lakes detected with ICESat altimetry in Antarctica²². There are no measurements of ice inflow for the interval between the formation of the basin and our first DEM. We therefore assume that the volume of the surface basin observed in May 2012 is equal to the volume of water lost during the 2011 subglacial lake outburst event. It is likely that the original volume of the basin was even larger. We calculate the minimum discharge rate by dividing the volume of the basin by the maximum time interval that elapses (21 days). We find that the subglacial lake emptied at a rate of about 215 m³ s^{–1}, which is more than three times the average flow rate of the River Thames in London and is also larger than drainage rates observed from Antarctic subglacial lakes. The subglacial drainage path for this flood can be estimated from hydrologic gradients²³, although confidence in the result is limited by the large uncertainties in the bed topography and the potential of an open conduit system²⁴. Hydraulic potential based on bed topography²⁵ and the InSAR DEM²¹ indicates

¹Earth and Atmospheric Sciences, Cornell University, Ithaca, New York 14853, USA. ²Department of Geological Sciences, University of North Carolina, Chapel Hill, North Carolina 27599, USA. ³Polar Geospatial Center, University of Minnesota, Saint Paul, Minnesota 55108, USA. ⁴School of Earth Sciences, Ohio State University, Columbus, Ohio 43210, USA. ⁵Lamont-Doherty Earth Observatory of Columbia University, Palisades, New York 10964, USA.

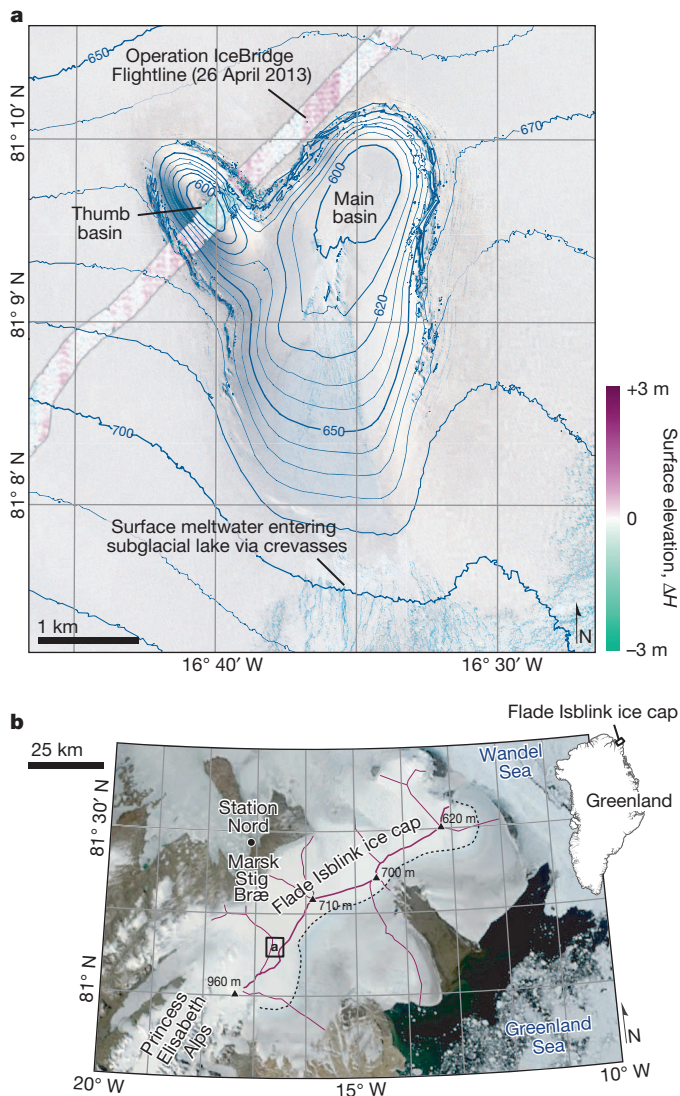


Figure 1 | Surface basin, Flade Isblink ice cap, northeast Greenland.

a, Composite satellite image of 70-m-deep basin. Supraglacial meltwater flowing northwards from the ice divide is being intercepted by marginal crevasses and routed to a subglacial lake. Scarring within the main basin is from an earlier supraglacial meltwater incursion. A NASA IceBridge airborne laser altimeter (lidar) profile³⁰ acquired on 26 April 2013 is colour-coded to show differences in elevation (ΔH) between the lidar and a corrected WorldView-1 satellite DEM collected 9 days later, on 5 May 2013 (see scale on right). 10-m contours of ice surface show the ellipsoidal elevation from WorldView-1 DEM acquired on 17 May 2012. Background image is from WorldView-2 on 14 August 2012. **b**, Regional setting. Purple lines show the ice divides derived from ref. 21. The dotted line shows the surface melt limit on background MODIS imagery from 11 July 2012. The inset shows location within Greenland and the box shows location of panel **a**. WorldView-2 imagery, copyright 2012, DigitalGlobe, Inc.; MODIS imagery courtesy of NASA.

that the floodwaters from the subglacial lake exited the ice cap to the north, beneath the Marsk Stig Bræ marine-terminating grounded outlet glacier²¹, west of Station Nord (Fig. 1).

We tracked the evolution of both the collapse basin and the associated surface drainage networks through the following two melt seasons using stereo pairs of half-metre-resolution WorldView satellite imagery over a 22-month period between May 2012 and March 2014 (for imagery and DEM details and uncertainties see Extended Data Table 1). Spatially varying elevation gains of between 2 m and 5 m occur in the main basin (Fig. 2) early in the first melt season following the collapse. We attribute these early season elevation gains (before 17 July 2012) to

drifting snow collecting in the basin^{18,26} and to ice inflow in response to the formation of the basin²⁷. On 13 July 2012, the north-flowing supraglacial meltwater network again formed close to the ice divide, ~5 km to the southwest of the 'mitten'. The supraglacial streams from this network disappear down crevasses on the southern margin of the main basin by 19 July 2012 (see Fig. 1 and Extended Data Fig. 3). Scars on the basin floor indicate that the supraglacial streams occasionally bypassed the crevasses and flowed into the body of the main basin (Fig. 1 and Extended Data Fig. 3). The north-flowing supraglacial streams continued to feed into the marginal crevasses through to 17 August 2012, the final high-resolution non-stereo image available for 2012. Meltwater cannot be seen in MODIS imagery on 30 August 2012.

During the two-week period of 31 July to 14 August 2012, meltwater drained into marginal crevasses and the floor of the main basin rose about 6 m (Fig. 3). This elevation gain is far faster than the 8 m yr^{-1} (0.02 m per day) observed over subglacial Lake Whillans in West Antarctica²⁸. The rapid uplift at the Flade Isblink ice cap ($>0.40 \text{ m per day}$) is the result of vertical motion localized in the basin floor (Fig. 2). This rapid uplift rate occurs only when surface meltwater enters the crevasses (Fig. 3). The uplift extends across the entire basin and is faster than other uplift documented during the evolution of the Flade Isblink ice cap basin. From observations of supraglacial meltwater entering crevasses and late-melt season uplift of the basin floor, we infer that supraglacial meltwater recharges the subglacial lake. From 17 May 2012 to 14 August 2012 the floor of the main basin rose a total of ~15 m.

The 'thumb' basin evolved differently from the main basin during the 2012 melt season. In early July 2012 a melt lake formed within the smaller basin. The surface of this lake rose ~5 m by 27 July 2012. While rapid uplift began in the main basin over the next four days, the ~140,000-m³ lake in the 'thumb' sub-basin catastrophically drained, leaving behind a 3–4-m-high pile of ice rubble (Fig. 2). This type of block uplift is seen at other rapidly draining supraglacial lakes around Greenland⁸.

Over the first winter (between mid-August 2012 and mid-April 2013), the main basin developed a pronounced topographic bulge. The floor of the basin rose an additional 15 m for a total of $>30 \text{ m}$ basin floor uplift relative to the basin elevation in May 2012 (Fig. 2). Between 17 May 2012 and 5 May 2013 the volume of the surface basin decreased by $46.5(0.6) \times 10^6 \text{ m}^3$. (The parentheses indicate 2σ uncertainties, the 95% confidence interval.) Ice inflow accounted for about a third of this volume change, $15.2(1.2) \times 10^6 \text{ m}^3$, while a numerical weather model¹⁸ provides snow accumulation of $1.7(0.3) \times 10^6 \text{ m}^3$ over this period. The majority of the volume loss, $29.6(1.1) \times 10^6 \text{ m}^3$ [total basin volume reduction – (ice inflow + snowfall volume)], is due to the uplift of the basin floor, which in turn is caused by the influx of supraglacial water into the subglacial lake. The numerical weather model¹⁸ predicts a remarkably similar amount of meltwater production, $31.8(6.4) \times 10^6 \text{ m}^3$ within the 57.2-km² catchment area that feeds into the crevasses around the collapse basin. The model predicts the end of surface meltwater production on 14 August 2012, while our observations show meltwater flowing for at least three more days.

The northerly flowing supraglacial drainage network was absent during the cooler summer of 2013. Localized streams and pools occasionally formed on the floor of the main basin. Over the 2013 melt season, the floor of both sub-basins rose ~4 m. Over the 2013–2014 winter (August 2013–March 2014) the basin floor rose about 4 m more and the basin lost a further $16.8(0.4) \times 10^6 \text{ m}^3$ of volume. The surface lowering around the periphery of the 'mitten' (Fig. 2d and Extended Data Fig. 4), which is probably partly due to ice flow into the basin, provides about half of this volume change, $7.2(1.3) \times 10^6 \text{ m}^3$. Snow accumulation estimates are not available for this time period. Other causes of the surface lowering adjacent to the basin may be the elastic response to the ice rising within the basin and to melting occurring around the edges of the subglacial lake during the injection of surface meltwater. Snow accumulation and ice flowing into the basin provide spatially varying, relatively low rates of surface uplift that are overprinted by rapid

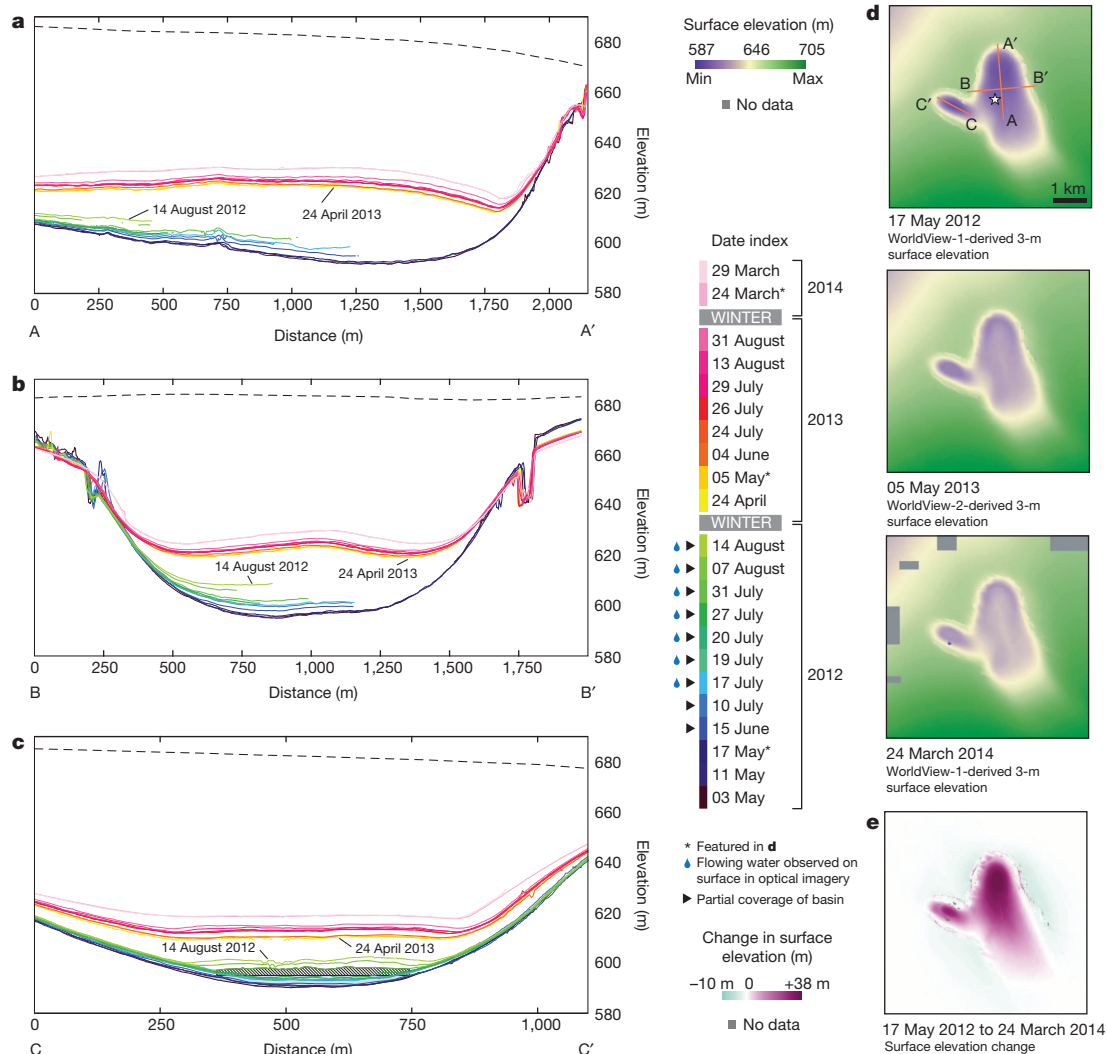


Figure 2 | Elevation profiles through time. The dotted black lines in **a** to **c** show the pre-collapse ice surface elevation modified from ref. 21. **a**, Repeat elevation profile A to A' (shown in **d**). Profile colour-indexed by time of satellite DEM acquisition (scale on right for **a** to **c**). Maximum uplift of around 38 m is seen at about 1,250 m from the start of the profile. The basin floor moves upwards through time, in part owing to the recharge of the subglacial lake. Uncertainties in elevation are typically ~ 0.5 m. **b**, Repeat elevation profile B to B' (shown in **d**). Maximum uplift of about 33 m occurs at about 1,000 m from the start of the profile. **c**, Repeat elevation profile C to C' (shown in **d**). Blue-grey background shading indicates the elevation gain due to formation of

supraglacial lake and hatched shading indicates single block uplift episode in response to drainage of a supraglacial lake. **d**, Examples of 3 m posting along-track stereo satellite-derived DEMs showing elevation gains over time in the basin. Star indicates location of elevation changes shown in Fig. 3. **e**, Change in surface elevation in part caused by recharge of subglacial lake between 17 May 2012 and 24 March 2014. Maximum displacement is ~ 38 m in the northern part of the main basin. Small amounts of subsidence occur around the periphery of the basin owing to ice flow into the basin, melting at the boundary of the subglacial lake and the elastic response of the surrounding ice to the basin floor rising.

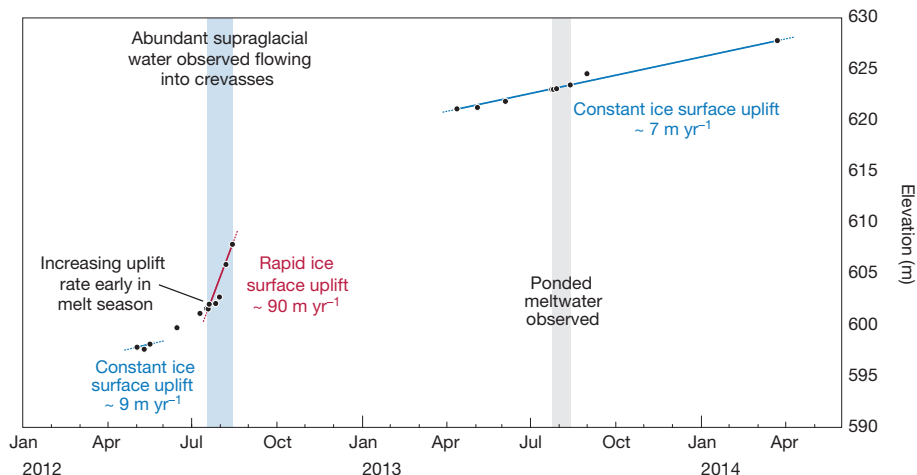


Figure 3 | Basin uplift rates through time.

Average ice surface elevation changes for a portion of the main basin. Location of uplift indicated by star in Fig. 2d. Uplift caused by snow accumulation and ice influx over the ~ 2 -yr study period is overprinted by a short period of rapid uplift when supraglacial water enters nearby crevasses. During 2013 this supraglacial drainage was absent and rates of uplift remain fairly constant. The timing and pattern of the short-term accelerated uplift occurs everywhere throughout the basin, although the absolute rates of uplift vary. Uncertainties on elevation measurements (black dots) are ~ 0.5 m.

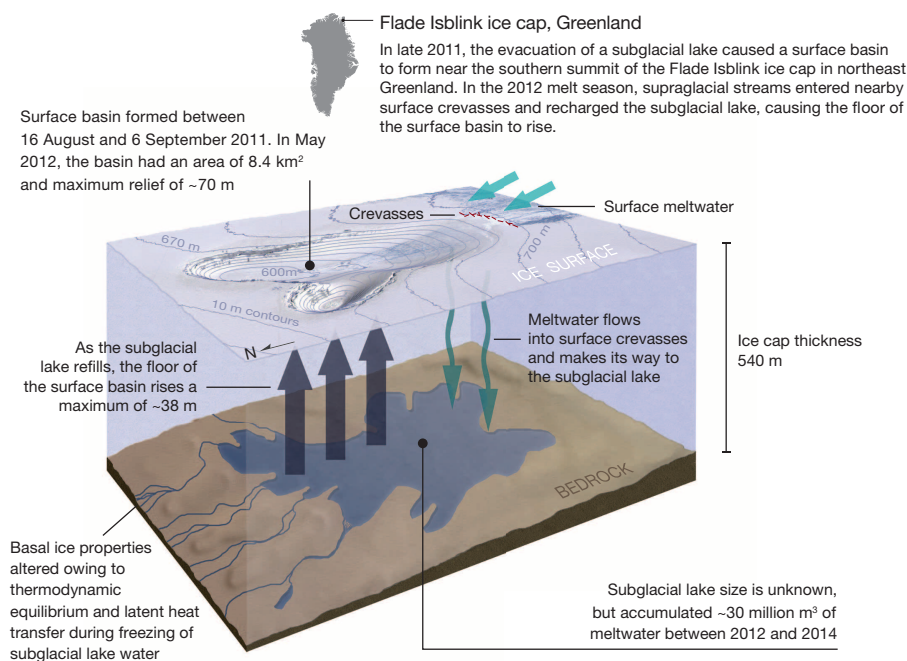


Figure 4 | Conceptual model of subglacial lake recharge from supraglacial streams at the Flade Isblink ice cap. Graphic is not to scale. WorldView-2 surface imagery, copyright 2012, DigitalGlobe, Inc. Inset shows location on Greenland.

elevation gains that occur when meltwater flows into nearby crevasses (Fig. 3).

Over about two years, the ice surface in the basin rose by up to 38 m and the basin lost a total volume of $67.3(0.4) \times 10^6 \text{ m}^3$. About half of this volume, $32.9(1.4) \times 10^6 \text{ m}^3$ is due to ice flow into the basin, while the remaining half, $34.4(1.5) \times 10^6 \text{ m}^3$, is composed of volume gains at the subglacial lake and the accumulation of snow. Our observations provide the first evidence of the direct recharge of a subglacial lake from surface meltwater (Figs 3 and 4). These results demonstrate that meltwater can be trapped and stored at the base of an ice cap. The subglacial environment in western Greenland is thought to react quickly to the transfer of surface meltwater to the bed. Our study indicates that in Greenland, and by inference at other ice sheets, this may not always be the case and that surface meltwater may be trapped for long periods. Supraglacial meltwater may fill some of the 400 or more subglacial basins predicted²⁹ to lie beneath Greenland. If these basins are filled from supraglacial melt, the basal ice in the vicinity of these lakes is likely to be warmer¹¹ and to have a lower viscosity than expected¹³. Basal warming combined with sporadic lake floods could modulate and possibly accelerate downstream ice flow, with implications for the mass balance of Greenland. The storage of surface meltwater in subglacial lakes provides a new mechanism for transferring the impacts of a warming atmosphere and increased surface melting to the critical base of the ice sheet.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 8 May; accepted 22 October 2014.

Published online 21 January 2015.

1. Iken, A. & Bindenschadler, R. Combined measurements of subglacial water pressure and surface velocity of Findelengletscher, Switzerland: conclusions about drainage system and sliding mechanism. *J. Glaciol.* **32**, 101–119 (1986).
2. Zwally, H. J. *et al.* Surface melt-induced acceleration of Greenland Ice-Sheet flow. *Science* **297**, 218–222 (2002).
3. Schoof, C. Ice-sheet acceleration driven by melt supply variability. *Nature* **468**, 803–806 (2010).
4. Stearns, L. A., Smith, B. E. & Hamilton, G. S. Increased flow speed on a large East Antarctic outlet glacier caused by subglacial floods. *Nature Geosci.* **1**, 827–831 (2008).
5. Bartholomew, I. *et al.* Short-term variability in Greenland Ice Sheet motion forced by time-varying meltwater drainage: implications for the relationship between subglacial drainage system behavior and ice velocity. *J. Geophys. Res.* **117**, F03002 (2012).

6. Andrews, L. C. *et al.* Direct observations of evolving subglacial drainage beneath the Greenland Ice Sheet. *Nature* **514**, 80–83 (2014).
7. Joughin, I. *et al.* Seasonal speedup along the western flank of the Greenland ice sheet. *Science* **320**, 781–783 (2008).
8. Das, S. B. *et al.* Fracture propagation to the base of the Greenland ice sheet during supraglacial lake drainage. *Science* **320**, 778–781 (2008).
9. Joughin, I., Alley, R. B. & Holland, D. M. Ice-sheet response to oceanic forcing. *Science* **338**, 1172–1176 (2012).
10. Murray, T. *et al.* Ocean regulation hypothesis for glacier dynamics in southeast Greenland and implications for ice sheet mass changes. *J. Geophys. Res.* **115**, <http://dx.doi.org/10.1029/2009JF001522> (2010).
11. Phillips, T., Rajaram, H., Colgan, W., Steffen, K. & Abdalati, W. Evaluation of cryo-hydrologic warming as an explanation for increased ice velocities in the wet snow zone, Sermeq Avannarleq, West Greenland. *J. Geophys. Res.* **118**, 1241–1256 (2013).
12. Thoma, M., Grosfeld, K., Mayer, C. & Pattyn, F. Ice-flow sensitivity to boundary processes: a coupled model study in the Vostok Subglacial Lake area, Antarctica. *Ann. Glaciol.* **53**, 173–180 (2012).
13. Bell, R. E. *et al.* Deformation, warming and softening of Greenland's ice by refreezing meltwater. *Nature Geosci.* **7**, 497–502 (2014).
14. Rennermalm, A. K. *et al.* Evidence of meltwater retention within the Greenland ice sheet. *Cryosphere* **7**, 1433–1445 (2013).
15. Palmer, S. J. *et al.* Greenland subglacial lakes detected by radar. *Geophys. Res. Lett.* **40**, 6154–6159 (2013).
16. Wolovick, M. J., Bell, R. E., Creyts, T. T. & Frearson, N. Identification and control of subglacial water networks under Dome A, Antarctica. *J. Geophys. Res.* **118**, 1–15 (2013).
17. Bell, R. E. The role of subglacial water in ice-sheet mass balance. *Nature Geosci.* **1**, 297–304 (2008).
18. Tedesco, M., Fettweis, X., Alexander, P., Green, G. & Datta, T. *MAR Greenland Outputs 1953–2013* (City University of New York Digital Archive, 2014).
19. Rogozhina, I. *et al.* Effects of uncertainties in the geothermal heat flux distribution on the Greenland Ice Sheet: An assessment of existing heat flow models. *J. Geophys. Res.* **117**, F02025 (2012).
20. Glasser, N. F. & Siebert, M. J. Calculating basal temperatures in ice sheets: an Excel spreadsheet method. *Earth Surf. Process. Landf.* **27**, 673–680 (2002).
21. Palmer, S. J., Shepherd, A., Sundal, A., Rinne, E. & Nienow, P. InSAR observations of ice elevation and velocity fluctuations at the Flade Isblink ice cap, eastern North Greenland. *J. Geophys. Res.* **115**, F04037 (2010).
22. Smith, B. E., Fricker, H. A., Joughin, I. R. & Tulaczyk, S. An inventory of active subglacial lakes in Antarctica detected by ICESat (2003–2008). *J. Glaciol.* **55**, 573–595 (2009).
23. Shreve, R. L. Movement of water in glaciers. *J. Glaciol.* **11**, 205–214 (1972).
24. Gulley, J. D. *et al.* The effect of discrete recharge by moulins and heterogeneity in flow-path efficiency at glacier beds on subglacial hydrology. *J. Glaciol.* **58**, 926–940 (2012).
25. Bamber, J. L. *et al.* A new bed elevation dataset for Greenland. *Cryosphere* **7**, 499–510 (2013).
26. Rinne, E. J. *et al.* On the recent elevation changes at the Flade Isblink Ice Cap, northern Greenland. *J. Geophys. Res.* **116**, F03024 (2011).
27. Aðalgeirsdóttir, G., Gudmundsson, G. H. & Björnsson, H. The response of a glacier to a surface disturbance: a case study on Vatnajökull ice cap, Iceland. *Ann. Glaciol.* **31**, 104–110 (2000).

28. Fricker, H. A., Scambos, T. A., Bindschadler, R. & Padman, L. An active subglacial water system in West Antarctica mapped from space. *Science* **315**, 1544–1548 (2007).
29. Livingstone, S. J., Clark, C. D., Woodward, J. & Kingslake, J. Potential subglacial lake locations and meltwater drainage pathways beneath the Antarctic and Greenland ice sheets. *Cryosphere* **7**, 1721–1740 (2013).
30. Krabill, W. B. *IceBridge ATM L2 Icessn Elevation, Slope, and Roughness. Version 2. Subset of data from 2013.04.26* <http://nsidc.org/data/icesat/> (NASA DAAC at the National Snow and Ice Data Center, 2014).

Acknowledgements The copyright for the satellite imagery is held by DigitalGlobe, Inc. We thank M. Studinger, T. Wagner, the NASA Operation IceBridge project team and the National Snow and Ice Data Center for the airborne radar and laser altimetry. We thank S. Palmer for providing the European Remote Sensing satellite InSAR DEM. We thank the University of North Carolina at Chapel Hill Research Computing group for providing computational resources that have contributed to these research results. We thank

M. Tedesco and X. Fettweis for help with MAR (Modele Atmospherique Regional) output. We thank T. Pavelsky, B. Mirus and J. Rich for comments and suggestions that improved the paper. This work was supported by US National Science Foundation grant number ARC-1111882. WorldView imagery was provided by the Polar Geospatial Center at the University of Minnesota, which is supported by grant ANT-1043681 from the US National Science Foundation.

Author Contributions M.J.W. designed the study, performed the analysis and wrote the paper. B.G.H. produced the diagrams and contributed to the paper. M.G.B. and R.E.B. contributed to the paper. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.J.W. (mike.willis@cornell.edu).

METHODS

Ice depths. Ice depths were measured with the University of Kansas Multichannel Coherent Radar Depth Sounder (MCoRDS) ice-penetrating radar³¹ flown by NASA's Operation IceBridge on 26 April 2013 (Extended Data Fig. 1).

DEMs. WorldView DEMs were derived using the Ames Stereo Pipeline³² applied to half-metre resolution, along-track satellite stereo image pairs acquired between 3 May 2012 and 29 March 2014. As few of our DEMs extend over bedrock, we cannot vertically co-register these DEMs using fixed ground control points. Instead we use the IceBridge Airborne Topographic Mapper (ATM) swath laser altimetry³⁰ from 26 April 2013 to correct a WorldView-1 DEM collected on 5 May 2013, nine days after the airborne mission. We remove the basin and nearby elevations from the ATM and WorldView elevation data and assume that the snow surface elevations more than 100 m away from the basin did not change over the interval between the ATM and the WorldView acquisitions. Differencing the surface heights indicates that the uncorrected satellite DEM is higher than the ATM data by 3.03 ± 0.25 m. We remove this offset, and then vertically co-register all our other DEMs to this 'corrected' WorldView DEM, normalizing our DEM heights.

Features align almost perfectly over multiple satellite images, so we do not horizontally co-register the DEMs or images. The 90% circular error of probability for the half-metre imagery is <4.0 m for Worldview-1 and better than 3.5 m for WorldView-2.

When differencing DEMs, initial outliers are removed by clipping elevation differences to ± 200 m. Remaining outliers are removed with an iterative 3-standard-deviation filter that finds the mean and standard deviation, clips the elevation differences to ± 3 standard deviations centred on the mean, then re-calculates the mean and standard deviation. We stop the iteration when the standard deviation changes by less than 2% between iterations (typically 3 or 4 iterations are needed).

Uncertainties. Standard deviations on the satellite derived DEMs are provided in Extended Data Table 1. ATM data³⁰ have uncertainties that are about ± 0.1 m (ref. 33). When performing calculations we add the ATM uncertainty to the DEM standard deviation in quadrature. In the worst case this provides an uncertainty of about ± 0.6 m, which if spread across the entire $\sim 8.4\text{-km}^2$ basin equates to a volume change of $\sim 5.0 \times 10^6 \text{ m}^3$. Additional uncertainty is from elevation changes that occur through time at the Flade Isblink ice cap. Elevation changes between 2004 and 2008 interpolated from ICESat tracks²⁶ near the location of the basin are about $+0.5$ m per year. We do not know whether this rate is representative of 2012–2014 surface elevation changes away from the basin.

To estimate the correlation length scale for volume uncertainties³⁴ we examine several pairs of DEMs that extend over $\sim 400 \text{ km}^2$ of bedrock to the north of the study area. Agreement is excellent, with elevation differences typically in the ± 1 m

range. We calculate how variance changes with increasing pixel sizes. The variance over bedrock pixels becomes independent at a length scale L of 150 m. For each volume calculation we use the formula $1.96U/\sqrt{N}$, where U is the "volume of uncertainty"³⁵, which is the combined DEM uncertainty multiplied by the total area A under consideration. N is the number of independent pixels found by dividing A by L^2 .

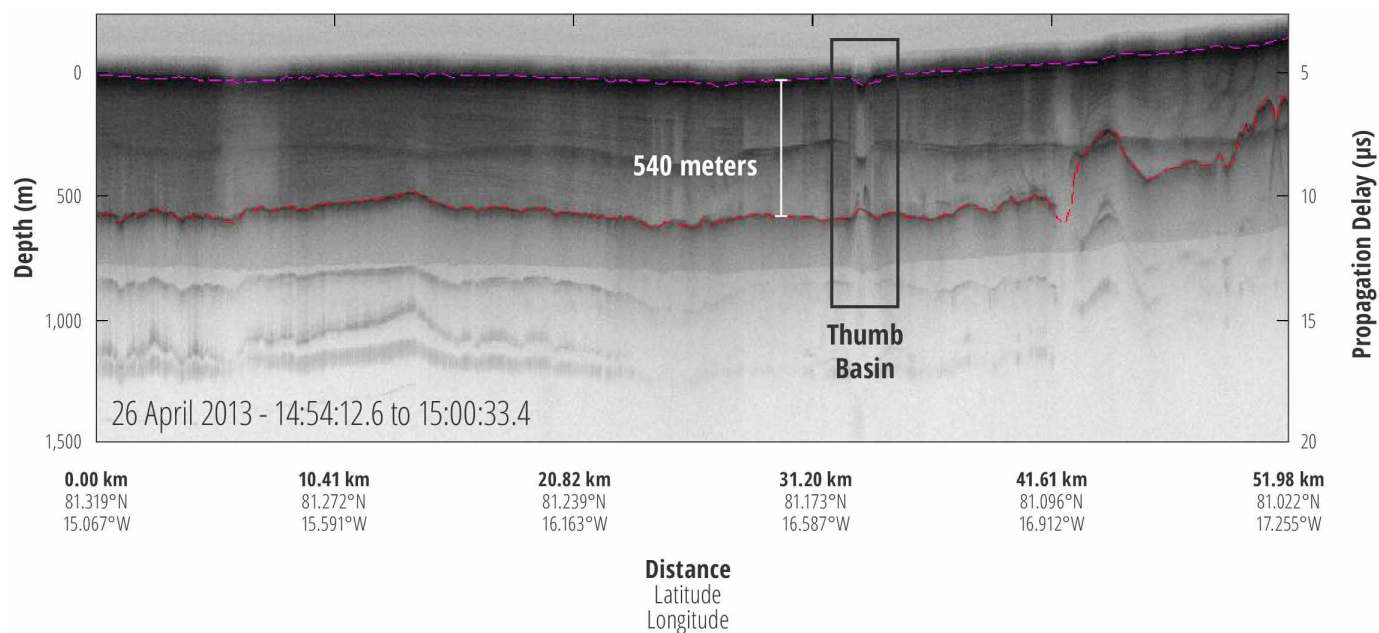
Ice inflow. Ice inflow into the basin is calculated by integrating ice surface elevation changes that occur between the lip of the basin out to a distance of 2 km (about four ice thicknesses) from the basin. We assume that the elevation falls occurring in this region correspond to ice flowing into the basin, although there could also be a small component of elevation change caused by deflation of the snow surface and melting at the subglacial lake.

Outliers are removed by clipping initial elevation differences to ± 200 m. Residual outliers are removed with an iterative 6-standard-deviation filter that finds the mean and standard deviation, clips the elevation differences to ± 6 standard deviations centred on the mean, then re-calculates the mean and standard deviation. We stop the iteration when the standard deviation changes by less than 2% between iterations (typically 1 or 2 iterations are needed). We choose a more lenient filter outside the basin than inside, to allow surface changes of up to 6 m adjacent to the basin into the calculation. The much larger area of small changes that occur farther away would otherwise dominate the calculation.

Snowfall. We use the cumulative daily surface mass balances from the MAR numerical weather model¹⁸ when available. The cumulative rate is converted to a thickness using an average snow density of 175 kg m^{-3} found from MAR for the period between 17 May 2012 and 5 May 2013. MAR uncertainties are set at 20%.

DEM availability. The DEMs used for this work are available at <http://www.pgc.umn.edu/elevation/stereo/flade>. The DEMs are produced using the Ames Stereo Pipeline (ASP), which is open source code available from NASA.

- Gogineni, P. *CRSIS Radar Depth Sounder Data* <http://data.cresis.ku.edu/> (2012).
- Moratto, Z. M., Broxton, M. J., Beyer, R. A., Lundy, M. & Husmann, K. Ames Stereo Pipeline, NASA's open source automated stereogrammetry software. *Lunar Planet. Sci. Conf.* **41** (2010).
- Krabill, W. B. *et al.* Aircraft laser altimetry measurement of elevation changes of the greenland ice sheet: technique and accuracy assessment. *J. Geodyn.* **34**, 357–376 (2002).
- Rolstad, C., Haug, T. & Denby, B. Spatially integrated geodetic glacier mass balance and its uncertainty based on geostatistical analysis: application to the western Svartisen ice cap, Norway. *J. Glaciol.* **55**, 666–680 (2009).
- Howat, I. M., Smith, B. E., Joughin, I. & Scambos, T. A. Rates of southeast Greenland ice volume loss from combined ICESat and ASTER observations. *Geophys. Res. Lett.* **35**, <http://dx.doi.org/10.1029/2008GL034496> (2008).



Extended Data Figure 1 | Operation IceBridge radar profile showing ice depths near the southern summit of Flade Isblink ice cap, northeast Greenland. Modified from a profile of the Multichannel Coherent Radar Depth Sounder (operated by the University of Kansas)³¹ over the ‘thumb’ basin collected on 26 April 2013. The NASA IceBridge flight line proceeds from approximately northeast to southwest and is shown in Fig. 1. The purple dotted line is the ice surface; the red dotted line is the ice/bed interface. The ‘thumb’

basin is at ~ 33 km along the flight line and shows an ice depth of ~ 540 m. The propagation delay is the time for the radar to travel from the aircraft to the ice and back; details of its conversion to depth can be found at the MCoRDS technical page: ftp://data.cresis.ku.edu/data/rds/rds_readme.pdf. The dielectric constant for ice (ϵ_r) used during the conversion from propagation delay to ice depth is set to 3.15 m.

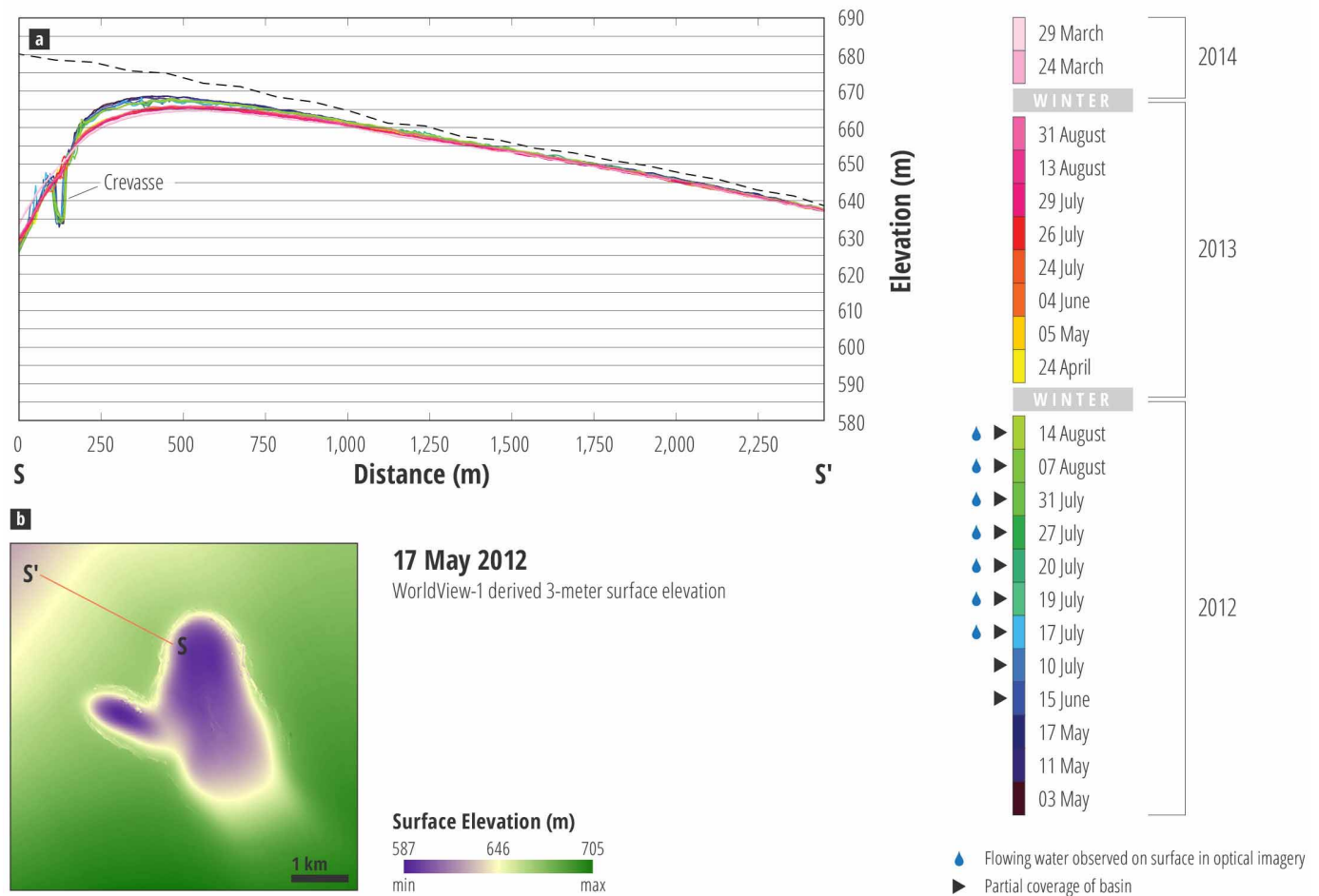


Extended Data Figure 2 | Surface collapse basin near the ice divide of Flade Isblink ice cap, northeast Greenland. Photographed from the north by M. Studinger, NASA Operation IceBridge, on 26 April 2013.



Extended Data Figure 3 | Collapse basin and northward-flowing supraglacial stream network near the ice divide of Flade Isblink ice cap, northeast Greenland. Supraglacial water disappears into crevasses at the edge of the basin. WorldView-2 multispectral image from 14 August 2012. Imagery copyright 2012, DigitalGlobe Inc.

Outside Basin Elevation Profiles



Extended Data Figure 4 | Elevation changes next to the basin. **a**, Repeat elevation profile S to S', adjacent to the basin, shown in **b**. Three to six metres of subsidence is seen within a kilometre of the rim of the basin. A crevasse observed in 2012 closes by 2013. The dotted black line is the pre-collapse ice

surface elevation modified from ref. 21. The profile is colour-indexed by time of satellite DEM acquisition (scale on right). Uncertainties in elevation are typically less than ~0.5 m. **b**, Location of profile on WorldView DEM.

Extended Data Table 1 | Details of satellite along-track stereo pairs

Date	Image-ID1	Image-ID2	Satellite	Mean (m)	Median (m)	Standard Deviation (m)	Number of Samples ($\times 10^6$)
03 MAY 2012	102001001ACA2000	102001001AB6D300	WV01	-3.54	-3.64	0.57	5.84
11 MAY 2012	102001001B090E00	102001001A49ED00	WV01	+0.04	+0.05	0.58	9.08
17 MAY 2012	102001001B25B800	102001001B6C0B00	WV01	-2.43	-2.45	0.44	18.67
15 JUN 2012	10300100192B8E00	1030010019346100	WV02	-3.11	-3.13	0.41	10.12
10 JUL 2012	103001001A3C4100	103001001AB78700	WV02	-4.72	-4.72	0.42	10.06
17 JUL 2012	1030010019553F00	103001001AA24600	WV02	-7.90	-7.91	0.38	9.25
19 JUL 2012	103001001B7CE000	10300100194C6600	WV02	-4.72	-4.73	0.37	9.94
20 JUL 2012	103001001B318B00	103001001A279800	WV02	-4.27	-4.27	0.41	8.65
27 JUL 2012	103001001A401E00	103001001AC27F00	WV02	-5.88	-5.88	0.37	9.91
31 JUL 2012	103001001A738F00	103001001AAC0B00	WV02	+2.81	+2.81	0.37	9.06
07 AUG 2012	103001001BCC9B00	103001001BC8BE00	WV02	+1.82	+1.80	0.36	9.89
14 AUG 2012	103001001B437500	103001001B07CC00	WV02	-4.97	-4.98	0.41	10.07
24 APR 2013	1030010022377600	10300100212B7B00	WV02	+1.70	+1.69	0.33	8.17
05 MAY 2013*	1030010022CC4100	10300100220CF000	WV02	+3.03	+3.03	0.25	0.30
04 JUN 2013	1030010023C32600	1030010023017F00	WV02	-1.58	-1.62	0.43	7.18
24 JUL 2013	1030010024218800	1030010024795500	WV02	-1.63	-1.65	0.31	18.43
26 JUL 2013	1030010025C35700	103001002551AB00	WV02	-1.77	-1.78	0.26	17.62
29 JUL 2013	1030010025215900	1030010025966A00	WV02	-3.74	-3.77	0.36	16.57
13 AUG 2013	10300100253F7500	10300100253F7500	WV02	-6.51	-6.53	0.38	14.88
31 AUG 2013	1030010026470C00	1030010027CA3A00	WV02	-0.66	-0.67	0.32	16.77
24 MAR 2014	102001002D5FEB00	102001002D4DD100	WV01	+4.97	+4.95	0.41	13.80
29 MAR 2014	103001002E6B0A00	103001002FB4F000	WV02	+7.88	+7.89	0.36	8.07

* The statistics for this DEM are calculated against a NASA Operation IceBridge laser altimetry swath flown over the area on 26 April 2013, nine days earlier³⁰.

Identification numbers are DigitalGlobe imagery codes. The images used are from along-track WorldView-1 and -2 half-metre panchromatic stereo imagery (WV01 and WV02). Mean, median and the standard deviation of elevation differences are computed by subtracting the 5 May 2013 DEM from the DEM in question for distances more than 100 m from the lip of the basin. The 5 May 2013 DEM is corrected to a NASA ATM laser altimetry swath collected on 26 April 2013 (ref. 30). The number of 3-m-resolution pixels used in the calculation of mean, median and standard deviation is given in the final column.

RBM3 mediates structural plasticity and protective effects of cooling in neurodegeneration

Diego Peretti¹, Amandine Bastide¹, Helois Radford¹, Nicholas Verity¹, Colin Molloy¹, Maria Guerra Martin¹, Julie A. Moreno¹, Joern R. Steinert¹, Tim Smith¹, David Dinsdale¹, Anne E. Willis¹ & Giovanna R. Mallucci^{1,2}

In the healthy adult brain synapses are continuously remodelled through a process of elimination and formation known as structural plasticity¹. Reduction in synapse number is a consistent early feature of neurodegenerative diseases^{2,3}, suggesting deficient compensatory mechanisms. Although much is known about toxic processes leading to synaptic dysfunction and loss in these disorders^{2,3}, how synaptic regeneration is affected is unknown. In hibernating mammals, cooling induces loss of synaptic contacts, which are reformed on rewarming, a form of structural plasticity^{4,5}. We have found that similar changes occur in artificially cooled laboratory rodents. Cooling and hibernation also induce a number of cold-shock proteins in the brain, including the RNA binding protein, RBM3 (ref. 6). The relationship of such proteins to structural plasticity is unknown. Here we show that synapse regeneration is impaired in mouse models of neurodegenerative disease, in association with the failure to induce RBM3. In both prion-infected and 5XFAD (Alzheimer-type) mice⁷, the capacity to regenerate synapses after cooling declined in parallel with the loss of induction of RBM3. Enhanced expression of RBM3 in the hippocampus prevented this deficit and restored the capacity for synapse reassembly after cooling. RBM3 overexpression, achieved either by boosting endogenous levels through hypothermia before the loss of the RBM3 response or by lentiviral delivery, resulted in sustained synaptic protection in 5XFAD mice and throughout the

course of prion disease, preventing behavioural deficits and neuronal loss and significantly prolonging survival. In contrast, knockdown of RBM3 exacerbated synapse loss in both models and accelerated disease and prevented the neuroprotective effects of cooling. Thus, deficient synapse regeneration, mediated at least in part by failure of the RBM3 stress response, contributes to synapse loss throughout the course of neurodegenerative disease. The data support enhancing cold-shock pathways as potential protective therapies in neurodegenerative disorders.

We used the phenomenon of physiological structural plasticity seen in hibernating mammals to determine the capacity for synapse regeneration in mouse models of neurodegenerative disease. When they enter torpor, the neurons of hibernators undergo morphological changes including changes in spine morphology^{8,9} and/or changes in connectivity^{4,5}. These are rapidly reversed on regaining normal body temperature^{4,5,8–10}. We first established that the phenomenon of synapse dismantling and reassembly (structural plasticity) on artificial cooling and rewarming occurs in laboratory mice (Fig. 1a and Extended Data Fig. 1a). We then explored the capacity for structural plasticity after cooling in two mouse models of neurodegenerative disease: prion disease and the 5XFAD model of Alzheimer's disease⁷. We used tg37^{+/-} mice¹¹ infected with Rocky Mountain Laboratory (RML) prions used in our previous studies^{12–14}. These mice show substantial synapse loss from 7 weeks post-inoculation

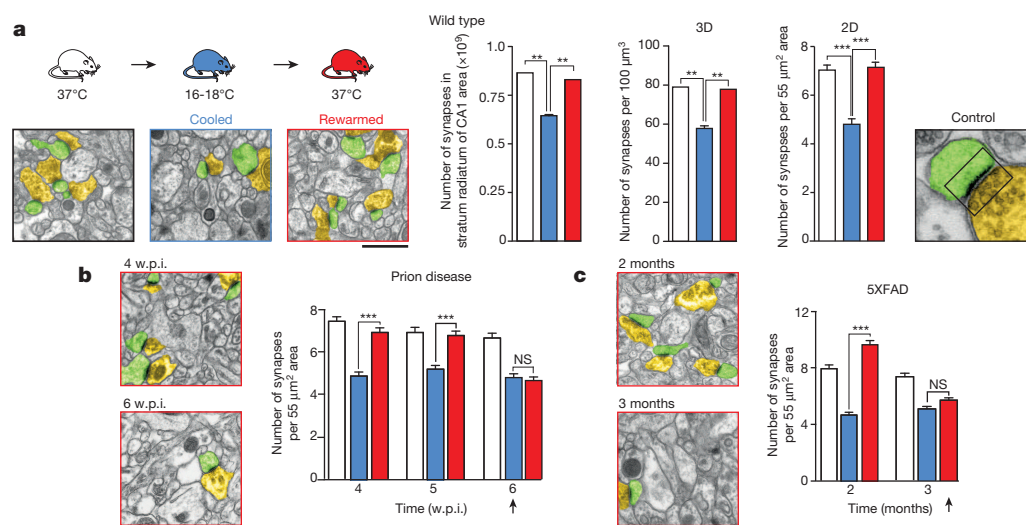


Figure 1 | The capacity for synaptic regeneration is lost early in neurodegenerative disease. **a**, Synapse numbers decline on cooling and recover on rewarming in wild-type mice, counted in both 3D and 2D. Representative electron micrographs (pseudo-coloured for ease of synapse identification: yellow, presynaptic; green, postsynaptic compartments) and bar charts showing quantification are shown for each experiment ($n = 4$ animals at 18 °C and $n = 2$ at 37 °C; 192 images from 2 mice per condition for 3D analyses;

93 images from 3 animals per condition, for 2D analyses). A typical tripartite synapse is shown at higher magnification. **b**, **c**, The same response is seen in prion-diseased mice (**b**) at 4 and 5 w.p.i. but this fails at 6 w.p.i. (arrow), and in 5XFAD (**c**) mice, where it fails at 3 months (arrow). *** $P < 0.0001$, ** $P < 0.01$; NS, not significant. Student's t -test; two tailed. All data in bar charts are mean \pm s.e.m. Scale bar, 1 μm . Source Data for all figures can be found in the Supplementary Tables.

¹Medical Research Council Toxicology Unit, Hodgkin Building, University of Leicester, Lancaster Road, Leicester LE1 9HN, UK. ²Department of Clinical Neurosciences, Clifford Allbutt Building, Cambridge Biomedical Campus, University of Cambridge, Cambridge CB2 0AH, UK.

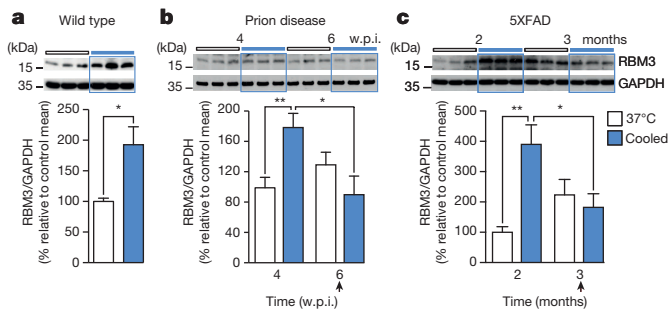


Figure 2 | Failure to induce RBM3 parallels lost capacity for synaptic recovery in neurodegenerative disease models. **a**, Cooling induces increased RBM3 levels in hippocampi of wild-type mice. **b**, **c**, The response fails at 6 w.p.i. in prion-infected mice (**b**) and at 3 months in 5XFAD mice (**c**) (arrows). Representative western blots are shown. Bar graphs show quantification of RBM3 levels relative to GAPDH, ($n = 6-11$ mice per time point; all experiments in triplicate) $**P < 0.01$, $*P < 0.05$, Mann-Whitney U -test in **a** and **c**, Student's t -test in **b**. All data are mean \pm s.e.m.

(w.p.i.)¹³; 5XFAD mice have synapse loss from 4 months, after which time learning deficits emerge⁷. We tested the capacity for structural plasticity using cooling early in the course of disease, before the onset of established synapse loss in both models: from 4 w.p.i. in prion-infected animals and from 2 months of age in 5XFAD mice.

All mice were cooled to 16–18 °C for 45 min, similar to core temperatures reached in small hibernators (deep hypothermia) using the biomolecule 5'-adenosine monophosphate (5'-AMP)¹⁵, after which they were allowed to slowly rewarm. Animals were euthanized at each stage of the cooling–rewarming process and synapses were counted in the CA1 region of hippocampus. Both synapse density and total synapse number significantly declined on cooling, but recovered on rewarming in wild-type mice, as measured using both three dimensional (3D)¹⁶ and two dimensional (2D) analyses¹³ (Fig. 1a). Neither brain volume nor synapse size changed on cooling and rewarming, excluding the possibility that changes in synapse density reflected changes in these parameters (Extended Data Fig. 1a). Thus, wild-type mice showed synaptic structural plasticity with reduction in synapse number on cooling and recovery on rewarming (Fig. 1a). This capacity for plasticity was also seen in both prion-infected and 5XFAD mice very early in the course of disease, at 4 and 5 w.p.i., and at 2 months of age, respectively (Fig. 1b, c). However, this capacity was lost by 6 w.p.i. in prion-diseased mice (Fig. 1b and Extended Data Fig. 1b) and at 3 months in 5XFAD mice (Fig. 1c and Extended Data Fig. 1c). Notably, impaired structural plasticity shortly preceded established decline in synapse number seen in prion-infected tg37^{+/-} mice at 7 w.p.i. (ref. 14), and in the 5XFAD mice from 4 months of age (see schematic, Extended Data Fig. 1d). The lost ability to reassemble synapses was not due to loss of synaptic proteins at this stage (Extended Data Fig. 2) nor to increased levels of disease-specific misfolded

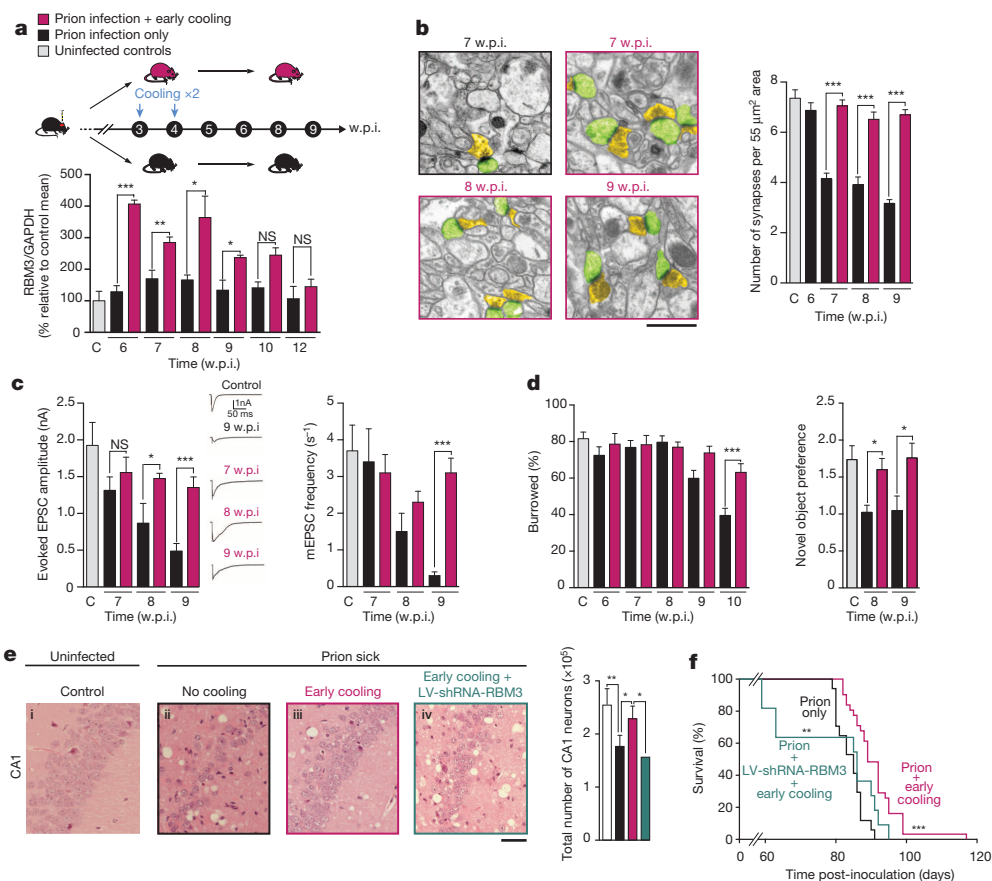


Figure 3 | Early cooling induces RBM3 overexpression and is neuroprotective in prion-infected mice. **a**, **b**, Cooling at 3 and 4 w.p.i. resulted in sustained high levels of RBM3 in hippocampus for several weeks ($n = 3-8$ mice per time point) causing marked protection (**b**) of synapse number at 7, 8 and 9 w.p.i. (62 images from 2 mice per time point). Scale bar, 1 μm . **c**, **d**, Early cooling maintained synaptic transmission ($n = 4-8$ cells from 2 mice per time point; representative raw traces of evoked EPSCs are shown) and prevented decline in burrowing behaviour and loss of novel object recognition memory (**d**), expressed as ratio of exploratory preference ($n \geq 10$ mice per group) in contrast to un-cooled mice. **e**, Haematoxylin and eosin stained sections show

striking reduction in hippocampal spongiosis and protection of CA1 neurons (bar chart) in cooled mice that is abolished by RBM3 knockdown ($n = 4-6$ animals per treatment, except LV-shRNA-RBM3: $n = 2$). Scale bar, 50 μm . One-way ANOVA, Brown–Forsythe test with Tukey's post hoc analysis for multiple comparisons was used in **d** and **e**. **f**, Early cooling significantly prolonged survival, but this was abolished by knockdown of RBM3 ($n = 31$ cooled mice; $n = 17$ not cooled; $n = 10$ cooled + RNAi of RBM3). Mann–Whitney U -test. $*P < 0.05$; $**P < 0.01$; $***P < 0.001$. Two-tailed Student's t -test was used unless otherwise stated. All data in bar charts are mean \pm s.e.m.

prion protein (PrP^{Sc}) in prion-infected mice, or of amyloid- β oligomers in 5XFAD mice induced by the cooling–rearming process (Extended Data Fig. 3).

In hibernation and hypothermia, global protein synthesis and cell metabolism are downregulated, but low temperature also induces a small subset of proteins known as cold-shock proteins that escape translational repression^{17,18}. Amongst these, RNA-binding motif protein 3 (RBM3) and cold-inducible RNA binding protein (CIRP, also known as CIRBP) are cold-shock proteins expressed at high levels in brain^{6,19}. We found strong induction of RBM3 by cooling in brains of wild-type mice and in mice with prion disease at 4 w.p.i. and 5XFAD mice at 2 months (Fig. 2). CIRP was not upregulated (Extended Data Fig. 4). However, both prion and 5XFAD mice lost the capacity to upregulate RBM3 after cooling at 6 w.p.i. (Fig. 2b) and at 3 months of age (Fig. 2c), respectively, in parallel with the lost ability to reassemble synapses after cooling at these time points (Fig. 1b, c). Therefore, we asked if induction of RBM3 expression drives synaptic recovery.

Therapeutic hypothermia is a powerful neuroprotectant in brain injury acting through multiple mechanisms, including enhanced gene expression driving regenerative processes enhancing synapse formation (see ref. 17 for a review). RBM3 has been implicated in protection against cell death in various *in vitro* models of cooling and neuroprotection^{20,21}, albeit in conditions of mild hypothermia ($\sim 32^{\circ}\text{C}$). It is known to increase local protein synthesis at dendrites¹⁹ and global protein synthesis through ribosomal subunit binding and/or microRNA biogenesis²². The neuroprotective effects of hypothermia on neurodegenerative disease are unknown, however. Given that the capacity for structural plasticity correlated with induction of RBM3, we asked if raising endogenous RBM3 levels through early therapeutic cooling would restore failed synaptic plasticity. In wild-type mice, a single episode of cooling to $16\text{--}18^{\circ}\text{C}$ raised

RBM3 levels in brain for up to 3 days (see Fig. 2a and Extended Data Fig. 4c), suggesting the response is sustained for some time after the cold stress. Animals were cooled twice: at 3 w.p.i. and again at 4 w.p.i., resulting in a sustained several-fold increase in RBM3 expression up to 6 weeks later, declining to baseline levels at 12 w.p.i., at the terminal stage of disease (Fig. 3a and Extended Data Fig. 5). Control mice were infected with prions but were not cooled. Early cooling and associated increased RBM3 expression protected against synapse loss in prion disease at 7, 8 and 9 w.p.i. (Fig. 3b), restored synaptic transmission (Fig. 3c) and prevented behavioural deficits, maintaining burrowing behaviours and novel object recognition memory (Fig. 3d and Extended Data Fig. 6a). There was also marked neuronal protection in the hippocampus (Fig. 3e, compare subpanels ii and iii), even in mice succumbing to prion infection, which is ultimately overwhelming due to other toxic effects¹³. Most remarkably, early cooling significantly increased survival in prion-infected mice (91 ± 7 days in cooled mice vs 84 ± 4 days for uncooled mice; $P = 0.0002$). Indeed, one animal survived 117 days post-infection, nearly a 50% increase in life expectancy (Fig. 3f). Mice cooled later in prion disease, at 5 and 6 w.p.i., when the RBM3 induction response is lost (see Fig. 2b), did not show increased survival (Extended Data Fig. 7). As predicted, RBM3 knockdown by lentivirally mediated RNA interference (RNAi) in the hippocampus abolished the protective effects of early cooling on CA1 pyramidal neurons and spongiform change (Fig. 3e, subpanel iv), on object recognition memory (Extended Data Fig. 6b, c), and on survival (Fig. 3f). As before, misfolded PrP levels were not affected by cooling (Extended Data Fig. 8). In therapeutic human hypothermia, temperatures of $\sim 34^{\circ}\text{C}$ are used, similar to those of hibernating large mammals such as bears, which are known to induce similar transcriptional changes in RBM3 (ref. 6). Therefore, these physiological changes in small rodents at $16\text{--}18^{\circ}\text{C}$ may well be relevant in therapeutic

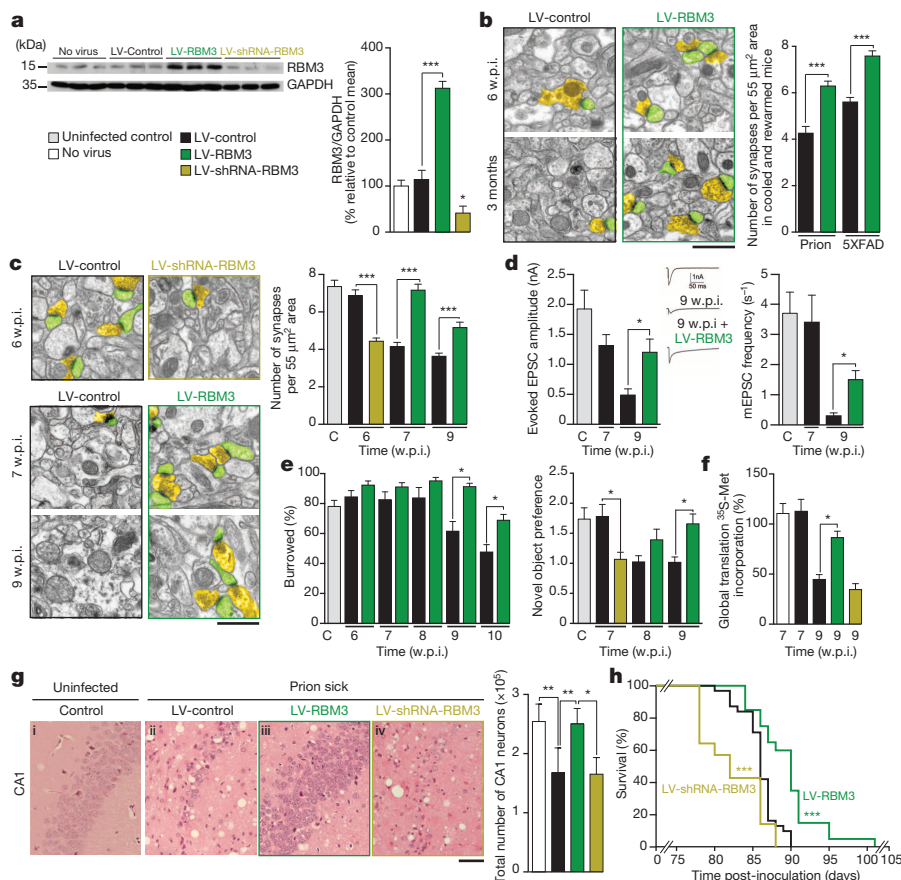


Figure 4 | Lentivirally mediated overexpression of RBM3 restores structural synaptic plasticity and is neuroprotective in neurodegenerative disease. **a**, LV-RBM3 produces high levels of expression in hippocampus; LV-shRNA-RBM3 reduces endogenous levels. Representative western blot and bar graph, where $n = 6$ mice per treatment, repeated in triplicate. **b**, **c**, LV-RBM3 rescues the deficit in structural plasticity in prion-infected and 5XFAD mice at 6 w.p.i. and 3 months of age, respectively (93 images from 3 mice per time point) and produced lasting protection of synapse number in prion-infected mice at 7 and 9 w.p.i. (**b**), whereas LV-shRNA-RBM3 accelerated synapse loss (**c**) (62 images from 2 mice per time point). **d**, **e**, LV-RBM3 resulted in significant recovery in synaptic transmission (**d**) (left panel, inset: representative raw traces of evoked EPSCs are shown; $n = 6$ cells from 2 mice per time point) and protected against behavioural decline (**e**), whereas RNAi of RBM3 accelerated memory loss and burrowing deficits (**e**) ($n \geq 10$ mice per group). **f**–**h**, RBM3 overexpression increases global protein synthesis rates in hippocampal slices of prion-infected mice at 9 w.p.i. (**f**) ($n = 4\text{--}6$ mice), is neuroprotective (**g**) ($n = 3\text{--}6$ mice) and significantly lengthened survival (**h**) ($n = 20$ mice), whereas knockdown accelerated disease ($n = 14$ mice) compared to LV-control treated mice ($n = 31$ mice), Mann–Whitney U -test. $*P < 0.05$, $**P < 0.01$, $***P < 0.001$. Student's t -test, two tails except in **e**, **f** and **g** which used one-way ANOVA, Brown–Forsythe test with Tukey's post hoc analysis for multiple comparisons. All data represent means \pm s.e.m. Scale bar, $1 \mu\text{m}$ in **b** and **c** and $50 \mu\text{m}$ in **g**.

human hypothermia²³. Cooling of mice to higher core temperature of 26–28 °C, was similarly protective in prion disease, extending survival (Extended Data Fig. 9).

We next asked if RBM3 overexpression alone, in the absence of cooling, was similarly neuroprotective. We overexpressed, or knocked-down, RBM3 in both hippocampi of mice by stereotaxic injection of lentiviruses LV-RBM3 and LV-shRNA-RBM3, respectively. LV-RBM3 produced a threefold increase in RBM3 levels compared to controls up to 8 weeks post-injection; whereas knockdown by LV-shRNA-RBM3 reduced RBM3 levels to 30% of control levels (Fig. 4a). LV-RBM3 treatment, but not LV-control, rescued the early deficit in synapse reassembly in both prion-infected and 5XFAD mice at 6 w.p.i. and 3 months, respectively (Fig. 4b). Furthermore, LV-RBM3 was associated with marked, sustained neuroprotective effects in prion-infected mice: preventing synapse loss (Fig. 4c), synaptic transmission decline (Fig. 4d) and memory and behavioural impairments (Fig. 4e). *In vitro*, RBM3 has been shown to promote translation²². Increased global protein synthesis rates are profoundly neuroprotective, rescuing synapse number in prion disease^{13,14}. We found that RBM3 overexpression rescued levels of global translation, whereas RBM3 knockdown further reduced them, in prion-infected mice at 9 w.p.i. (Fig. 4f) suggesting that this action of RBM3 along with preferential translation of specific RBM3-bound mRNAs, contributes to the synapse regeneration process. LV-RBM3 treatment reduced prion neuropathology and prevented neuronal loss (compare subpanels ii and iii in Fig. 4g) and significantly extended survival of prion-infected animals (Fig. 4h). This was not associated with changes in levels of PrP^{Sc}, which were not affected by overexpression of RBM3 (Extended Data Fig. 8). Knockdown of RBM3, in contrast, accelerated synapse loss and memory and behavioural deficits (Fig. 4c–e), accelerating neuronal loss (Fig. 4g, subpanel iv) and significantly shortening survival (Fig. 4h). 5XFAD mice do not allow similar examination of long-term effects of RBM3 overexpression as evolution of deficits and neuronal loss takes many months, and life expectancy is normal. However, RNAi of RBM3 accelerated onset of synaptic loss in 5XFAD mice, which was now seen at 3 months (Extended Data Fig. 10a), suggesting that RBM3 has a long-term protective role in structural plasticity in these mice also. RBM3 knockdown also reduced synapse number and novel object memory in wild-type mice (Extended Data Fig. 10b), thus it is likely to be involved in synaptic maintenance under normal physiological conditions.

In conclusion, we have shown that early synapse loss in mouse models of neurodegenerative disease results, at least in part, from defective synaptic repair processes associated with failure to induce the cold-shock RNA-binding protein, RBM3. This results in impaired synaptic reassembly after cooling, but also appears to be important in the context of protecting against ongoing synaptic toxicity during disease, and in synaptic maintenance in wild-type mice. Our data suggest that further understanding the mechanisms of action of cold-shock proteins such as RBM3 may yield insights into endogenous repair processes and bring new therapeutic targets for neuroprotection in neurodegenerative disease.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 25 September 2013; accepted 5 December 2014.

Published online 14 January 2015.

- Holtmaat, A. & Svoboda, K. Experience-dependent structural synaptic plasticity in the mammalian brain. *Nature Rev. Neurosci.* **10**, 647–658 (2009).

- Selkoe, D. J. Alzheimer's disease is a synaptic failure. *Science* **298**, 789–791 (2002).
- Mallucci, G. R. Prion neurodegeneration: starts and stops at the synapse. *Prion* **3**, 195–201 (2009).
- Magariños, A. M., McEwen, B. S., Saboureau, M. & Pevet, P. Rapid and reversible changes in intrahippocampal connectivity during the course of hibernation in European hamsters. *Proc. Natl Acad. Sci. USA* **103**, 18775–18780 (2006).
- Popov, V. I. & Bocharova, L. S. Hibernation-induced structural changes in synaptic contacts between mossy fibres and hippocampal pyramidal neurons. *Neuroscience* **48**, 53–62 (1992).
- Williams, D. R. et al. Seasonally hibernating phenotype assessed through transcript screening. *Physiol Genomics* **24**, 13–22 (2005).
- Oakley, H. et al. Intraneuronal β -amyloid aggregates, neurodegeneration, and neuron loss in transgenic mice with five familial Alzheimer's disease mutations: potential factors in amyloid plaque formation. *J. Neurosci.* **26**, 10129–10140 (2006).
- Popov, V. I. et al. Reversible reduction in dendritic spines in CA1 of rat and ground squirrel subjected to hypothermia-normothermia in vivo: A three-dimensional electron microscope study. *Neuroscience* **149**, 549–560 (2007).
- Ruediger, J. et al. Dynamics in the ultrastructure of asymmetric axospinous synapses in the frontal cortex of hibernating European ground squirrels (*Spermophilus citellus*). *Synapse* **61**, 343–352 (2007).
- von der Ohe, C. G., Garner, C. C., Darian-Smith, C. & Heller, H. C. Synaptic protein dynamics in hibernation. *J. Neurosci.* **27**, 84–92 (2007).
- Mallucci, G. R. et al. Post-natal knockout of prion protein alters hippocampal CA1 properties, but does not result in neurodegeneration. *EMBO J.* **21**, 202–210 (2002).
- Mallucci, G. R. et al. Targeting cellular prion protein reverses early cognitive deficits and neurophysiological dysfunction in prion-infected mice. *Neuron* **53**, 325–335 (2007).
- Moreno, J. A. et al. Sustained translational repression by eIF2 α -P mediates prion neurodegeneration. *Nature* **485**, 507–511 (2012).
- Moreno, J. A. et al. Oral treatment targeting the unfolded protein response prevents neurodegeneration and clinical disease in prion-infected mice. *Sci. Transl. Med.* **5**, 206ra138 (2013).
- Zhang, J., Kaasik, K., Blackburn, M. R. & Lee, C. C. Constant darkness is a circadian metabolic signal in mammals. *Nature* **439**, 340–343 (2006).
- West, M. J. Counting and measuring ultrastructural features of biological samples. *Cold Spring Harb. Protoc.* <http://dx.doi.org/10.1101/pdb.top071886> (2013).
- Yenari, M. A. & Han, H. S. Neuroprotective mechanisms of hypothermia in brain ischaemia. *Nature Rev. Neurosci.* **13**, 267–278 (2012).
- Danno, S. et al. Increased transcript level of RBM3, a member of the glycine-rich RNA-binding protein family, in human cells in response to cold stress. *Biochem. Biophys. Res. Commun.* **236**, 804–807 (1997).
- Smart, F. et al. Two isoforms of the cold-inducible mRNA-binding protein RBM3 localize to dendrites and promote translation. *J. Neurochem.* **101**, 1367–1379 (2007).
- Chip, S. et al. The RNA-binding protein RBM3 is involved in hypothermia induced neuroprotection. *Neurobiol. Dis.* **43**, 388–396 (2011).
- Tong, G. et al. Effects of moderate and deep hypothermia on RNA-binding proteins RBM3 and CIRP expressions in murine hippocampal brain slices. *Brain Res.* **1504**, 74–84 (2013).
- Dresios, J. et al. Cold stress-induced protein Rbm3 binds 60S ribosomal subunits, alters microRNA levels, and enhances global protein synthesis. *Proc. Natl Acad. Sci. USA* **102**, 1865–1870 (2005).
- Lee, C. C. Is human hibernation possible? *Annu. Rev. Med.* **59**, 177–186 (2008).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank J. Edwards, J. McWilliam and D. Read of the MRC Toxicology Unit; members of University of Leicester Department of Biological Services staff; and D. Morrison, University of Cambridge, for technical assistance. We thank J. Skepper, University of Cambridge, for advice about stereological procedures. This work was funded by the Medical Research Council, UK.

Author Contributions D.P. performed most experimental procedures and analyses. A.B. analysed cold-shock proteins, N.V. and M.G.M. carried out prion inoculations and stereotaxic injections, C.M. performed behavioural tests, J.A.M. and H.R. carried out histological analyses, J.R.S. performed neurophysiological procedures, M.G.M. and T.S. performed ultramicrotomy and processed samples for electron microscopy, which was analysed by D.D. A.E.W. provided expertise on cold-shock and protein expression. G.R.M. conceived and directed the project. D.P. and G.R.M. wrote the paper. All authors contributed to discussion and analysis of data and to the final draft of the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to G.R.M. (gm522@cam.ac.uk).

METHODS

Animals. All animal work conformed to UK regulations and institutional guidelines and were performed under UK Home Office guidelines.

Prion infection of mice. The 3-week-old tg37^{+/−} mice¹¹ were inoculated intracerebrally into the right parietal lobe with 30 µl of 1% brain homogenate of Chandler/RML (Rocky Mountain Laboratories) prions under general anaesthetic, as described²⁴. Animals were culled when they developed clinical signs of scrapie as defined in¹⁴. Control mice received 1% normal brain homogenate.

5XFAD mice. Founder 5XFAD mice were obtained from the Jackson Laboratory (Ba Harbour, ME, USA). The 5XFAD mice have the following five mutations: Swedish (K670N and M671L), Florida (I716V) and London (V717I) in human APP695 and human PSI cDNA (M146L and L286V) under the transcriptional control of the neuron-specific mouse *Thy-1* promoter⁷. Colonies were maintained by crossing hemizygous transgenic to wild-type littermates.

Induction of hypothermia. FVB, tg37^{+/−} and 5XFAD mice weighing ≥ 20 g were cooled using 5'-AMP as described^{15,25}, with slight modifications. Freshly prepared 5'-AMP (Sigma) was injected intraperitoneally (0.7 mg per g). Control mice were injected with saline. Mice were maintained at room temperature until core body temperature decreased to 25 °C (approximately 60 min). Subsequently, mice were transferred to a refrigerator (5 °C) and core body temperature lowered to 16–18 °C for 45 min. Mice recovered normal body temperature at room temperature conditions. Cooled samples were collected at the end of the 16–18 °C period and rewarmed samples as stated elsewhere in the text.

Electron microscopy data acquisition and analysis of synapse number. Male mice were used to avoid the effects on synapse number of the oestrus cycle. Brains were perfusion fixed with 2% glutaraldehyde + 2% paraformaldehyde in 0.1 M sodium cacodylate buffer (final pH 7.3). Slices (300 µm) were prepared using a vibrating blade microtome (Leica Microsystems, Milton Keynes, UK). These slices were post-fixed in 1% osmium tetroxide + 1% potassium ferrocyanide, stained *en bloc* with 5% uranyl acetate and embedded in epoxy resin (TAAB Laboratories Equipment Ltd, Aldermaston, UK), as described²⁶.

For random sampling and calculation of synapse density, estimation of total synapse number and measures of volume of tissue using the stereology disector method, the following procedures were used, as described previously¹⁶. Fixed and stained slices were flat embedded as described²⁷. The slices were mounted for microtomy, a semi-thin (1 µm) section was cut from the upper surface of each slice and the area of the stratum radiatum (sr) measured using the Cavalieri estimator (points grid), as described¹⁶. These results were then used to estimate the volume of the stratum radiatum and the systematic random selection of regions for electron microscopy, as described¹⁶. Briefly, the grid has intersections every 50 pixels (at a scale that 1 pixel is 1 µm). Each intersection was labelled with a dot. The dots were numbered in each of the sections generated per hippocampus. These numbers were used for measuring the area of each section and generate the data about the volume applying the following formula: $V_{srCA1} = \Sigma PrCA1 \times Ap \times T$ (ref. 16). The total number of dots per hippocampus was divided by 6 (number of samples selected to obtain representative synapse variability from the whole hippocampus). The number obtained was used for the generation of a random number for sampling the first position and defined the interval for the subsequent 5 positions to be sampled. These regions were trimmed-down, mounted on aluminium pins and imaged by serial block-face scanning electron microscopy in a FEI Quanta FEG 250 electron microscope (FEI Ltd, Cambridge, UK) equipped with a '3View2XP' system (Gatan Ltd, Abingdon, UK). Images of 32 serial sections were generated in each of the six points. Serial sections of areas 51.4 µm² were recorded at $\times 20,000$ and an accelerating voltage of 3 kV, using spot size 3.0, at a working distance of 6.1 mm, and a dwell-time of 5 µs per pixel. The mean section thickness was estimated by sectioning only half of the original block face. The block was then re-embedded, sectioned orthogonally and the depth removed by '3View' was measured by transmission electron microscopy²⁸. The mean thickness of 3View sections was 86 nm. We analysed synapses in a volume stack of 88 µm³ (with an area of 33 µm², a measured section thickness of 0.086 µm and 31 sections). Synapses that had their first identifiable profile below the first section in the series were counted¹⁶. Synapses were identified, within a counting frame of 5.75 µm \times 5.75 µm, which follows the counting frame rules to avoid edge effect for the estimation of synapse numerical density. The total number of synapse was estimated with the synapse density and the volume of each hippocampus, as described¹⁶.

Synapse size was calculated by measuring mean synapse length, mean synapse area in the same serial sections used for estimations of synapse density and total synapse number, as described¹⁶.

For routine 2D analyses semi-thin (1 µm) sections were stained with toluidine blue and examined to select areas for ultramicrotomy. Ultrathin sections (~70 nm) were stained with lead citrate and examined, blind, in a Jeol 100-CXII electron microscope (JEOL (UK) Ltd, Welwyn Garden City, UK) equipped with a 'Megaview III' digital camera (Olympus Soft Imaging Solutions GmbH, Münster, Germany).

A series of images were recorded from the stratum radiatum, all at a distance of approximately 100 µm from the CA1 pyramidal layer to avoid the large dendritic profiles in the proximal area. 31 images, each encompassing an area of 55 µm², from each of two to three mice were used for scoring. For synapse quantification the following criteria were followed: the presence of an unambiguous postsynaptic density, a clear synaptic cleft, and three or more synaptic vesicles. An average of 300 synapses were counted per sample.

Quantification of numbers of neurons in CA1 region of hippocampus. For CA1 pyramidal layer volume analysis, whole hippocampus was cut and every 250 µm stained with haematoxylin and eosin in fixed sections of 5 µm thickness. The volume of the pyramidal layer was measured using the Cavalieri estimator (points grid). For neuron mean density slices were stained with NeuN and calculated within an unbiased virtual space. The total number of neurons in CA1 pyramidal layer was estimated with the neuron density and the volume of the hippocampus.

Immunoblotting. Protein samples were isolated from hippocampi using protein lysis buffer (50 mM Tris, 150 mM NaCl, 1% Triton X-100, 1% Na deoxycholate, 0.1% SDS and 125 mM sucrose) supplemented with Phos-STOP and protease inhibitors (Complete, Roche), followed by centrifugation and quantification. Protein levels were determined by resolving 20 µg of protein on SDS-polyacrylamide gel electrophoresis gels, transferred onto either nitrocellulose or PDVF membranes and incubated with primary antibodies. Synaptic proteins were detected using the following antibodies: SNAP-25, (1:10,000; catalogue number: ab5666, Abcam), VAMP2 (1:5,000; catalogue number: 104204, Synaptic Systems), NMDA-R1 (1:1,000; catalogue number: G8913, Sigma) and PSD95 (1:1,000; catalogue number: 04-1066, Millipore). Odyssey IRDye800 secondary antibodies (1:5,000; catalogue number: 926-32210/32211 LI-COR) were applied, visualized and quantified using Odyssey infrared imager (LI-COR; software version 3.0). Protein for PrP levels was determined using the primary antibody ICSM35 (1:10,000; catalogue number: 0130-03501, D-GEN). PrP^{Sc} was detected after Proteinase K digestion. Cold-shock protein levels were determined with antibodies CIRP (1:1,000; catalogue number: 10209-2-AP, Proteintech Group, Inc.) and RBM3 (1:500; catalogue number: 14363-1-AP, Proteintech Group, Inc.). Amyloid-β levels in 5XFAD mice were detected by 6E10 clone antibody (1:1,000; catalogue number: SIG-39320, Covance). Horseradish-peroxidase-conjugated secondary antibodies (1:10,000; DAKO) were applied and protein visualized using enhanced chemiluminescence (GE Healthcare) and quantified using ImageJ. An antibody against GAPDH (1:5,000; catalogue number: sc32233, Santa Cruz) was used to determine gel loading.

Lentiviruses. GenTarget (San Diego, CA, USA) generated lentiviral plasmids. The neuron-specific promoter CAMKII was used to drive RBM3; the H1 promoter was used for shRNA-RBM3 expression and scrambled sequence-shRNA. Viruses were injected stereotactically into the CA1 region of the hippocampus as described²⁹.

Mouse RBM3 isoform 2 (NM_001166410.1) overexpression was induced using the pLentiCAMKII (RBM3)Rsv (GFPBsd) plasmid. pLentiCAMKII (empty)Rsv (GFPBsd) was used as control. RBM3 down regulation was achieved by using pLentiH1shRNA(m RBM3) sequence number 2Rsv(GFPBsd). This plasmid contains the following shRNA-RBM3 sense, anti-sense and loop sequences (sequence number 2: 5'-GTTGATCATGCAGGAAAGTCTcgagAGACTTTCCTGCATGATCAAC-3').

pLentiH1shRNA (negative control)Rsv (GFPBsd) containing the sequence 5'-GTCTCCACGCGCAGTACATTT-3' was used as control. Lentiviral sequences and viral stocks were generated by GenTarget (San Diego, CA, USA). Virus titre was determined using FACS (BD FACS Calibur). Viruses were used with a final titre of 0.6×10^8 to 1.5×10^8 transducing units.

Stereotaxic injection. Under general anaesthetic, mice were injected with 5 µl of lentivirus per site into the CA1 region of the hippocampus. Mice were injected at 2 locations per hemisphere; at −2 mm and −2.7 mm posterior, ± 2 mm lateral and −2.2 mm ventral relative to bregma, using a 26 s-gauge needle and Hamilton syringe as described²⁹.

Burrowing assay. This was performed as described^{14,30}. Briefly, mice were placed in individual large plastic cages containing a clear Perspex tube, 20 cm long \times 6.8 cm diameter, filled with 140 g of normal food pellets. The weight of pellets remaining in the tube was measured after 2 h and the percentage burrowed calculated. Behavioural data were analysed using one-way ANOVA with Brown–Forsythe test and Tukey's post hoc test. For behavioural testing no formal randomization was needed or used. Experimenter was blind to group allocation during all experiments and when assessing outcome.

Novel object recognition memory. This was performed as described¹². Briefly, mice were tested in a black cylindrical arena (69 cm diameter) mounted with a 100 LED strip infrared light source and a high resolution day/night video camera (Sony). Mice were acclimatized to the arena 5 days before testing. During the learning phase, two identical objects were placed 15 cm from the sides of the arena. Each mouse was placed in the arena by an operator blind to the experimental group for two blocks of 10 min for exploration of the objects with an inter-trial interval of 10 min.

Two hours later, one of the objects was exchanged for a novel one, and the mouse was replaced in the arena for 5 min (test phase). The amount of time spent exploring all objects was tracked and measured for each animal using Ethovision software (Tracksys). All objects and the arena were cleansed thoroughly between trials to ensure the absence of olfactory cues. The amount of time spent exploring the novel object over the familiar object is expressed as a ratio, where a ratio of 1 reflects random exploration, and >1 reflects memory. Behavioural data were analysed using one-way ANOVA with Brown–Forsythe test and Tukey’s post hoc test. For behavioural testing no formal randomization was needed or used. Experimenter was blind to group allocation during all experiments and when assessing outcome.

Electrophysiology. Whole-cell recordings were made in acute hippocampal slices to measure synaptic transmission from identified CA1 neurons and recording performed as described³¹. In brief, neurons were voltage clamped using a Multiclamp 700B amplifier and pClamp 10.3 software (Molecular Devices) and EPSCs were evoked by stimulation with bipolar platinum electrode at 37 °C. Pipettes (2.5–3.5 M Ω) were filled with a solution containing (in mM): KCl 110, HEPES 40, EGTA 0.2, MgCl₂ 1, CaCl₂ 0.1; pH was adjusted to 7.2 with KOH. Neurons were visualized with $\times 60$ objective lenses on a Nikon FS600 microscope fitted with differential interference contrast optics. Four to eight cells were measured per mouse in at least two animals per experiment. Male mice were used to avoid effects of the oestrus cycle.

Hippocampal slice preparation and ³⁵S-methionine labelling. Slices were dissected in an oxygenated cold (2–5 °C) sucrose artificial cerebrospinal fluid (ACSF) containing (mm): 26mM NaHCO₃, 2.5mM KCl, 4 mM MgCl₂, 0.1 mM CaCl₂, and 250mM sucrose. Hippocampal slices were prepared using a tissue chopper (McIlwain). Slices were allowed to recover in normal ACSF buffer while being oxygenated at 37 °C for 1 h, then incubated with [³⁵S]-methionine label for 1 h, then homogenized¹⁴. Proteins were TCA precipitated and incorporation of radiolabel was measured by scintillation counting (Winspectral, Wallac).

Statistics. Statistical analyses were performed using Prism v5 software, using Student’s *t*-test for data sets with normal distribution and a single intervention; when the *F*-test to compare variances was significant, Mann–Whitney *U*-test was performed instead.

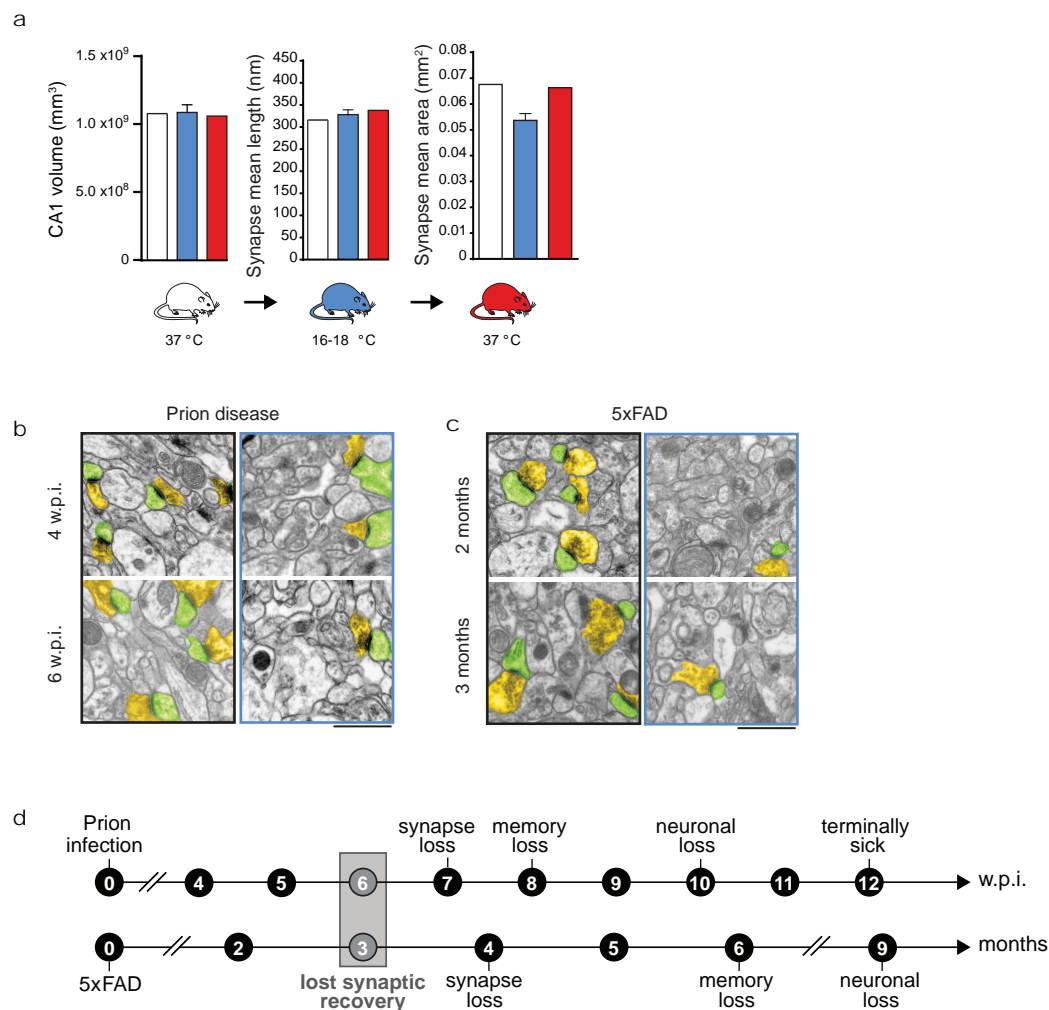
Behavioural data, neuronal counts and ³⁵S-met were analysed using one-way ANOVA and Tukey’s post hoc test for multiple variables.

For behavioural testing no formal randomization was needed or used. Experimenter was blind to group allocation during the experiments and when assessing outcome. **Statistical analyses for *in vivo* experiments.** Sample size estimation for induction of hypothermia for volume, synaptic density and estimation of total number of synapses was based in effect size of the preliminary experiment on dissector method (5.9511) and obtained with the free software G*Power version 3.1.9.2.

The software prediction shows that with a sample size of 6 animals for 2 conditions (control and cooled or cooled and rewarmed), the experiment has a 99.8% of chance of detecting a difference and avoid a type II error (β -error), with a 0.05% chance of a type I (α -error). Sample size estimation for novel object recognition experiment was established based on the effect size of 1.6161 from control and prion mice at 8 w.p.i. This parameter was applied in the following *F* tests calculation of power analysis with G*Power version 3.1.9.2.

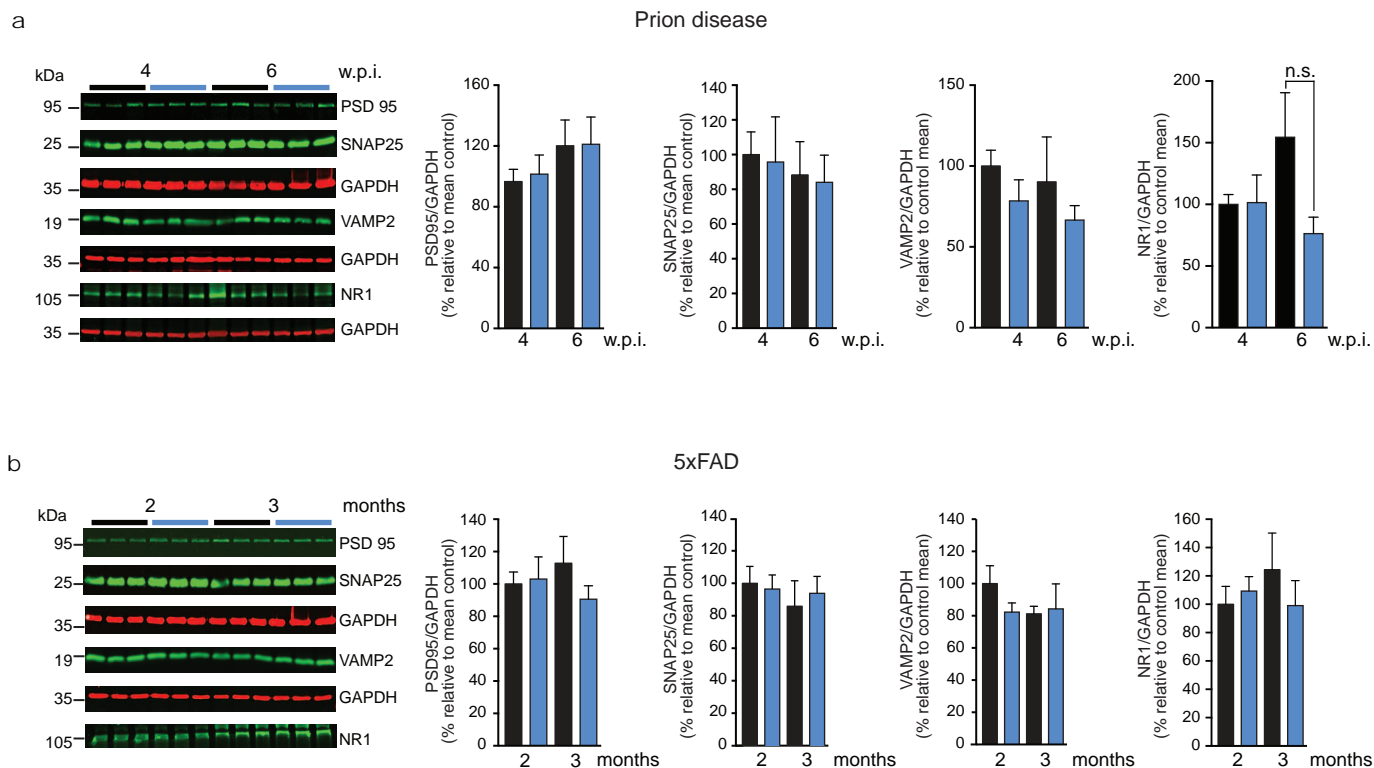
Similar analyses were performed for burrowing tests.

24. Mallucci, G. *et al.* Depleting neuronal PrP in prion infection prevents disease and reverses spongiosis. *Science* **302**, 871–874 (2003).
25. Daniels, I. S. *et al.* A role of erythrocytes in adenosine monophosphate initiation of hypometabolism in mammals. *J. Biol. Chem.* **285**, 20716–20723 (2010).
26. Deerinck, T. J. Enhancing serial block-face scanning electron microscopy to enable high resolution 3-D nanohistology of cells and tissues. *Microsc. Microanal.* **16**, 1138–1139 (2010).
27. Nguyen, K. B. & Pender, M. P. A simple technique for flat osmication and flat embedding of immunolabelled vibratome sections of the rat spinal cord for light and electron microscopy. *J. Neurosci. Methods* **65**, 51–54 (1996).
28. Yang, G. C. H. & Shea, S. M. Precise measurement of thickness of ultrathin sections by a ‘re-sectioned-section’ technique. *J. Microscopy* **103**, 385–392 (1975).
29. White, M. D. *et al.* Single treatment with RNAi against prion protein rescues early neuronal dysfunction and prolongs survival in mice with prion disease. *Proc. Natl Acad. Sci. USA* **105**, 10238–10243 (2008).
30. Cunningham, C. *et al.* Synaptic changes characterize early behavioural signs in the ME7 model of murine prion disease. *Eur. J. Neurosci.* **17**, 2147–2155 (2003).
31. Hausteiner, M. D. *et al.* Acute hyperbilirubinaemia induces presynaptic neurodegeneration at a central glutamatergic synapse. *J. Physiol.* **588**, 4683–4693 (2011).



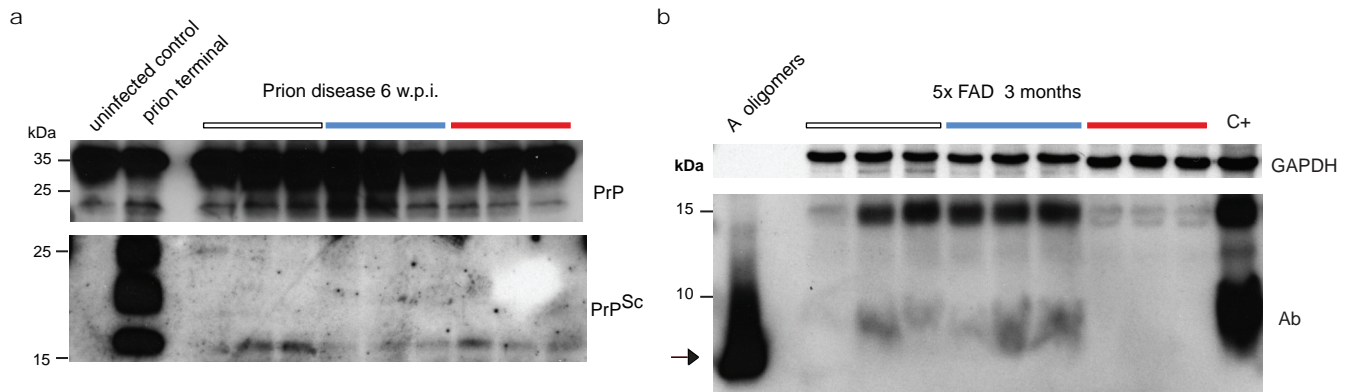
Extended Data Figure 1 | Stereological assessment of volume and synapse size to validate 2D assumption-based approaches for counting synapse density. **a**, CA1 volume and synapse mean length and area in the stratum radiatum remain essentially unchanged on cooling and rewarming in wild-type mice. Volume was measured using disector principle and synapse mean length and area determined in the same sections, as described¹, *n* values as reported for Fig. 1a. **b**, **c**, Representative electron micrographs (pseudo-coloured for

ease of synapse identification) for data not shown in Fig. 1b, c, from prion-infected mice at 4 and 6 w.p.i. (**b**) and for 5xFAD mice at 2 and 3 months (**c**) before cooling (black framed images) and cooled (blue framed images). **d**, Schematic showing lost capacity for structural plasticity precedes synapse loss and neuronal loss in both mouse models. Scale bar, 1 μ m. All data in bar charts are mean \pm s.e.m. Student's *t*-test, two tailed. Non-significant *P* values.



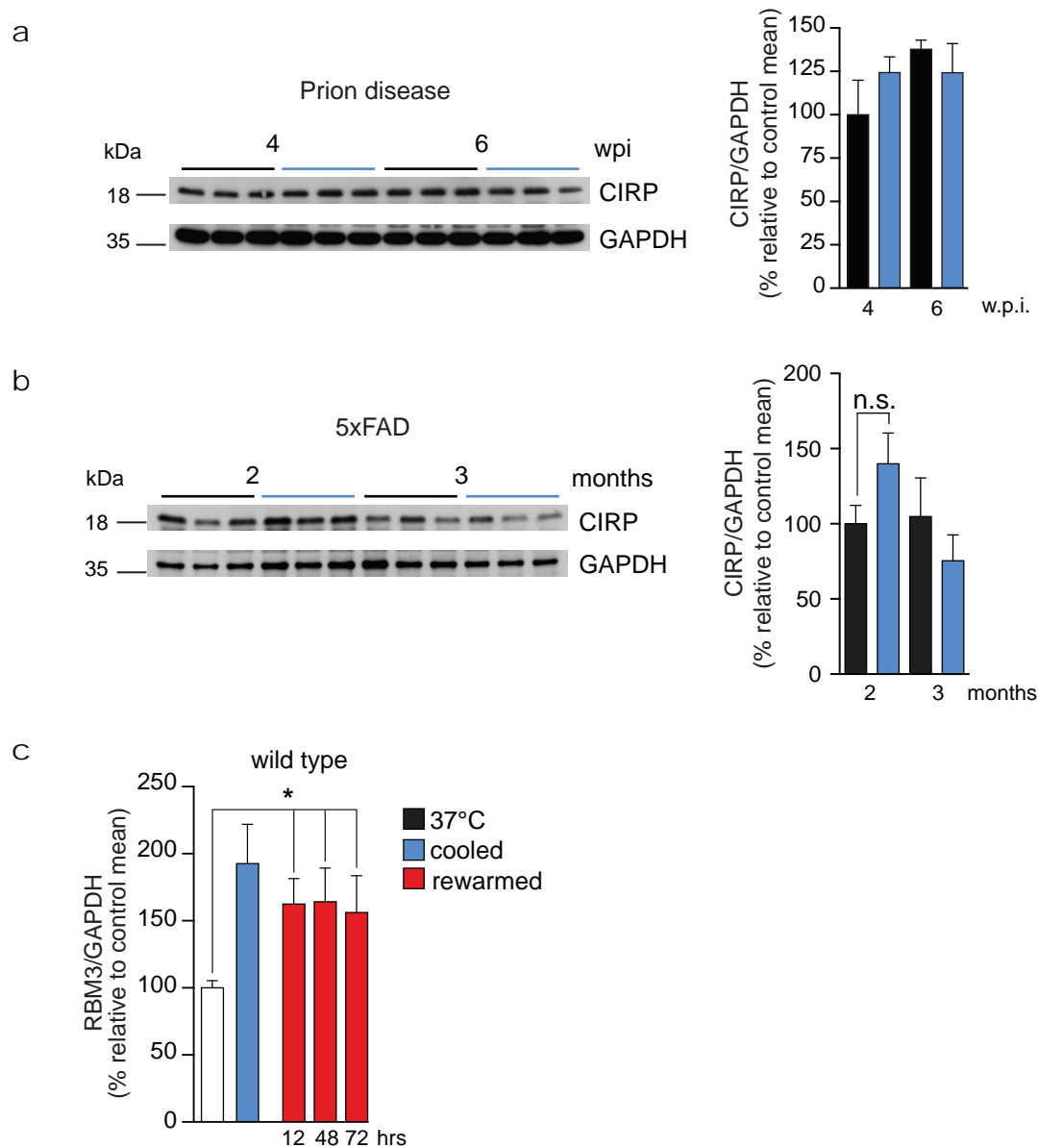
Extended Data Figure 2 | Synaptic protein levels during cooling–rewarming in prion and 5XFAD mice. **a**, Levels of presynaptic (SNAP25, VAMP2) and postsynaptic (PSD95, NR1) proteins do not change before (black bars) and after cooling to 16–18 °C (blue bars) in prion-infected mice at 4 and 6 w.p.i., **b**, 5XFAD mice at 2 and 3 months. Representative western blots

are shown for 3 mice per temperature and time point. Bar graphs show quantification of synaptic protein levels relative to GAPDH. All data represent means \pm s.e.m. ($n = 3–11$ mice per time point). Student's t -test, two tailed. n.s. = non-significant P values.



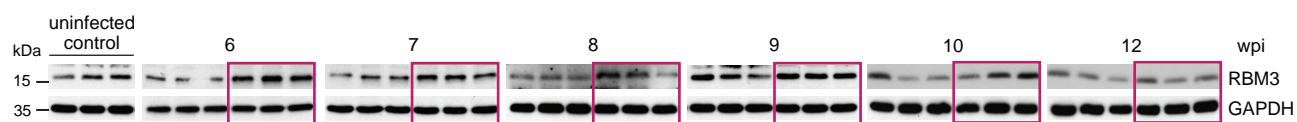
Extended Data Figure 3 | Cooling does not induce changes in PrP^{Sc} or amyloid- β levels. **a**, Levels of total PrP (upper blot) and PrP^{Sc} (lower blot) do not change notably before (white line), during (blue line) or after (red line) cooling to 16–18 °C in prion-infected mice. PrP^{Sc} is detected after digestion with proteinase K. Levels are undetectable by western blotting at 6 w.p.i., as

expected. **b**, Cooling does not change levels of amyloid- β oligomers in 5XFAD mice, arrow indicates amyloid- β monomers (lane 1, synthetic amyloid- β oligomers; last lane, one-year-old 5XFAD control (C+)). Representative western blots are shown for 3 mice per temperature and time point. Non-significant *P* values.



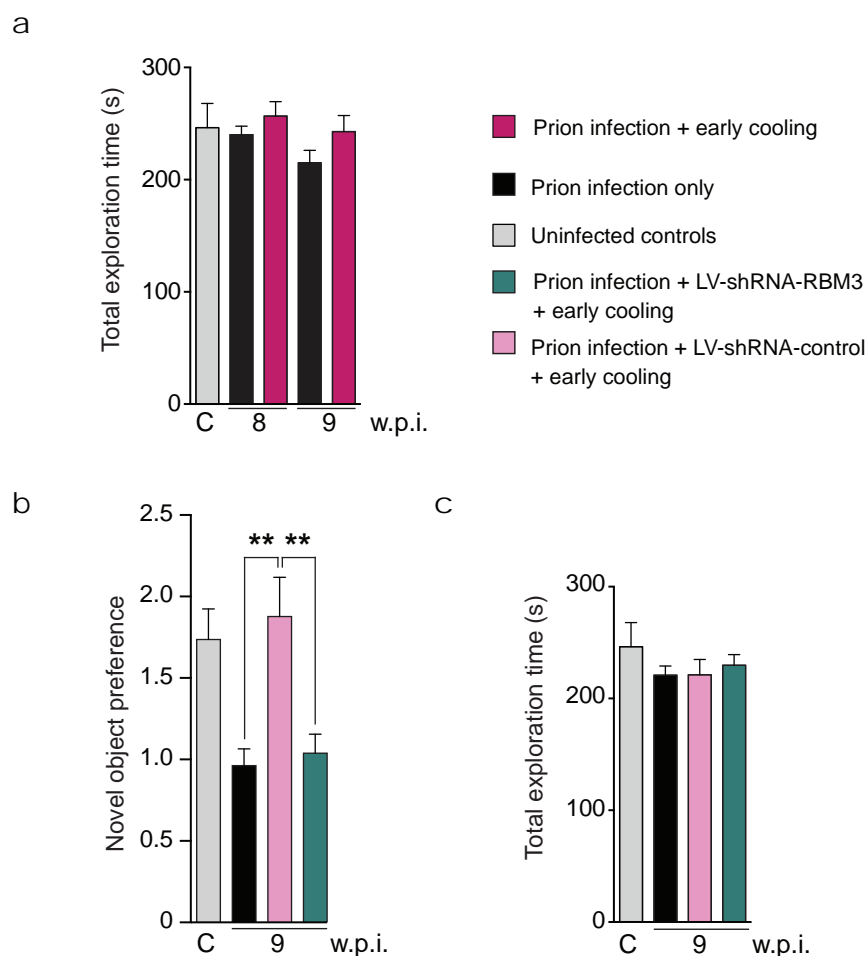
Extended Data Figure 4 | Cooling induces sustained increase in RBM3 levels but not in CIRP. **a, b,** Levels of CIRP do not change after cooling in prion-infected mice at 4 and 6 w.p.i. (**a**) or in 5xFAD mice at 2 and 3 months (**b**). Representative western blots are shown for 3 mice per temperature and time point. Bar graphs show quantification of CIRP levels relative to GAPDH. All data represent means \pm s.e.m. ($n = 6-9$ mice per time point). Student's

t -test, two tailed. n.s. = non-significant P values. **c,** Increased levels of RBM3 are sustained for at least 72 h after cooling in wild-type mice. Bar graph shows quantification of RBM3 against GAPDH in control (white bar), cooled (blue bar), and 12, 48 and 72 h recovery after cooling (red bars). All data represent means \pm s.e.m., ($n = 3-6$ mice per time points, $*P < 0.05$, Mann-Whitney U -test, two tailed).



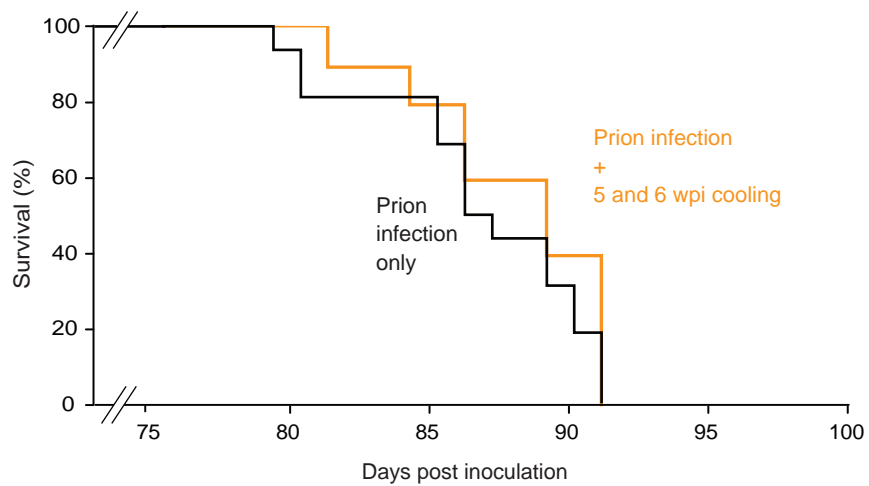
Extended Data Figure 5 | Early cooling induces sustained elevation of RBM3 levels. RBM3 levels remain high after cooling to 16–18 °C in prion-infected mice (magenta boxes) compared to control prion-infected mice. These

levels remained high up to 6 weeks later and declined at 12 w.p.i. Representative western blots are shown for 3 mice per time point.



Extended Data Figure 6 | Exploration time in exposure phase of novel object testing is normal in all groups and RBM3 knockdown abolishes improved memory after cooling. **a**, Exploratory behaviour measured in seconds is not different in mice with early cooling from prion-diseased mice and is not affected by the duration of disease (n as reported in Fig. 3d). **b**, **c**, Lentivirally mediated RNAi of RBM3 eliminates the protective effect of

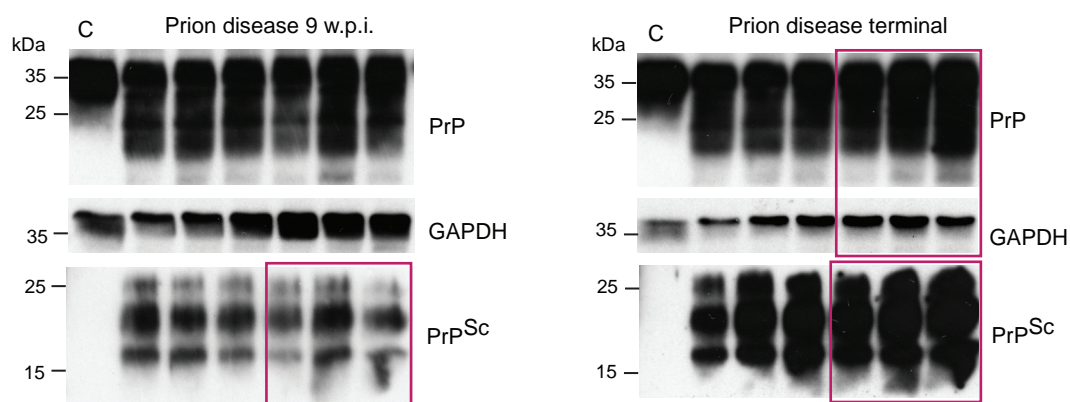
cooling on novel object memory impairment in prion disease (**b**) (dark green bar); but does not affect exploratory behaviour in training phase (**c**). All data represent means \pm s.e.m. Data analysed using one way ANOVA, Brown-Forsythe test with Tukey's post hoc analysis for multiple comparisons ($n = 11-16$ mice per time point, $**P < 0.01$).



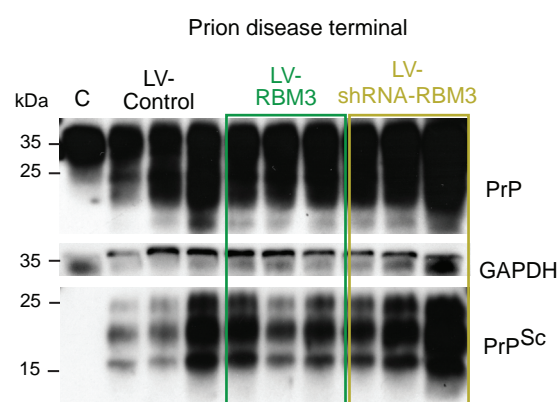
Extended Data Figure 7 | Induction of hypothermia at time point when RBM3 induction fails is not neuroprotective. Cooling at 5 and 6 w.p.i., when synaptic plasticity and RBM3 induction fails (see Fig. 1 and 2, main text), does

not increase survival in prion-infected mice. Kaplan–Meier survival plots for prion-infected mice (black line, no cooling; $n = 10$; orange line, mice cooled at 5 and 6 w.p.i., $n = 16$). Student's t -test, two tailed. Non-significant P values.

a

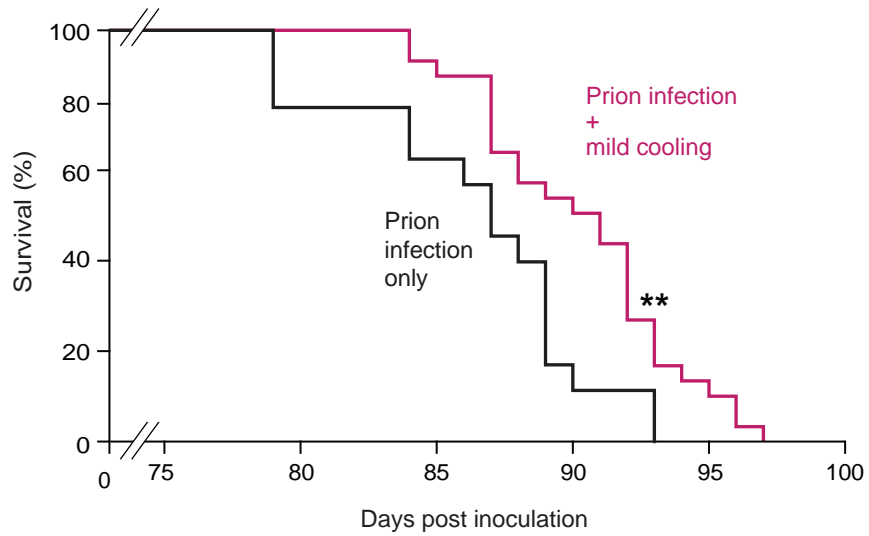


b

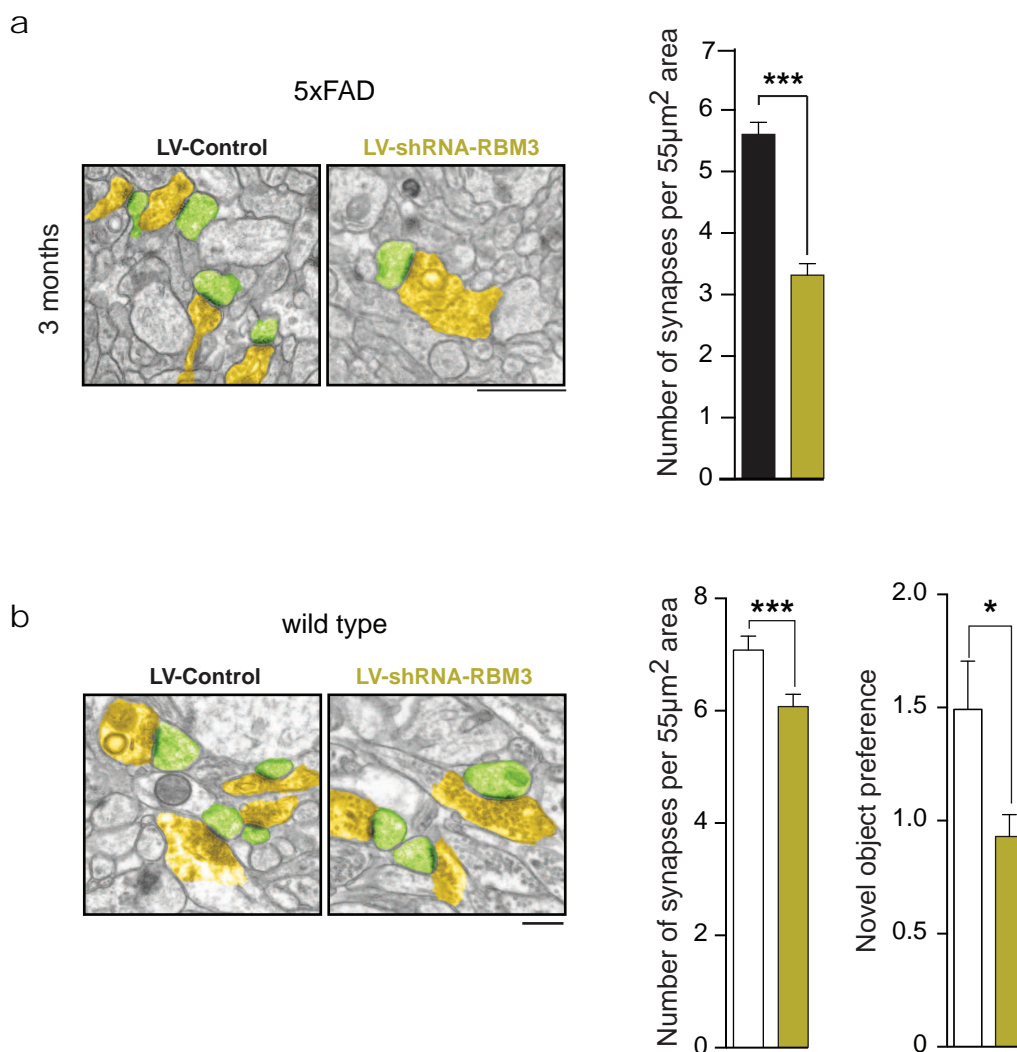


Extended Data Figure 8 | PrP^{Sc} levels remain unchanged in prion with overexpression of RBM3. a, b, In prion-infected mice total PrP and PrP^{Sc} levels do not alter after early cooling to 16–18 °C (a) (magenta boxes) or following treatment with LV-RBM3 (dark green) and LV-shRNA-RBM3

(pale green) (b). PrP and PrP^{Sc} levels tested in 9 w.p.i. and terminal mice. PrP^{Sc} is detected after digestion with proteinase K. Representative western blots are shown for 3 mice per temperature and time point, the lane marked C shows uninfected control mouse.



Extended Data Figure 9 | Mild hypothermia also extends survival in prion-infected mice. Kaplan–Meier plot showing that cooling to 26 °C at an early stage also significantly lengthens survival ($n = 27$ cooled vs $n = 16$ non-cooled mice); $**p < 0.01$, Student's t -test, two tailed.



Extended Data Figure 10 | RNAi of RBM3 downregulation accelerates impaired structural synaptic plasticity in the 5XFAD mouse model, and also reduced synapse number and function in wild-type mice. **a**, Impaired structural synaptic plasticity after cooling occurs in shRNA-RBM3 treated 5XFAD mice at 3 months. Representative electron micrographs are shown and are pseudo-coloured as in main text figures. Quantification shows significant

reduction in synapse number by RNAi of RBM3 ($n = 82$ – 93 images from 3 mice per time point, Student's t -test, two tailed). **b**, RBM3 knockdown reduces synapse number and novel object memory in wild-type mice ($n = 93$ images from 3 mice per time point, Student's t -test, two tailed $***P < 0.0001$; for novel object recognition task $n = 11$ mice, LV-shRNA-control and 10 mice, LV-shRNA-RBM3, Mann-Whitney U -test, $*P < 0.05$). Scale bar, 1 µm.

Convergent loss of PTEN leads to clinical resistance to a PI(3)K α inhibitor

Dejan Juric^{1*}, Pau Castel^{2*}, Malachi Griffith^{3,4,5}, Obi L. Griffith^{4,5,6}, Helen H. Won^{2,7}, Haley Ellis², Saya H. Ebbesen⁸, Benjamin J. Ainscough⁵, Avinash Ramu⁵, Gopa Iyer^{2,9}, Ronak H. Shah², Tiffany Huynh¹, Mari Mino-Kenudson¹, Dennis Sgroi¹, Steven Isakoff¹, Ashraf Thabet¹, Leila Elamine¹, David B. Solit^{2,9}, Scott W. Lowe^{8,10}, Cornelia Quadl¹¹, Malte Peters¹¹, Adnan Derti¹², Robert Schegel¹², Alan Huang¹², Elaine R. Mardis^{3,4,5,6}, Michael F. Berger^{2,7}, José Baselga^{2,13} & Maurizio Scaltriti²

Broad and deep tumour genome sequencing has shed new light on tumour heterogeneity and provided important insights into the evolution of metastases arising from different clones^{1,2}. There is an additional layer of complexity, in that tumour evolution may be influenced by selective pressure provided by therapy, in a similar fashion to that occurring in infectious diseases. Here we studied tumour genomic evolution in a patient (index patient) with metastatic breast cancer bearing an activating *PIK3CA* (phosphatidylinositol-4,5-bisphosphate 3-kinase, catalytic subunit α , PI(3)K α) mutation. The patient was treated with the PI(3)K α inhibitor BYL719, which achieved a lasting clinical response, but the patient eventually became resistant to this drug (emergence of lung metastases) and died shortly thereafter. A rapid autopsy was performed and material from a total of 14 metastatic sites was collected and sequenced. All metastatic lesions, when compared to the pre-treatment tumour, had a copy loss of *PTEN* (phosphatase and tensin homolog) and those lesions that became refractory to BYL719 had additional and different *PTEN* genetic alterations, resulting in the loss of PTEN expression. To put these results in context, we examined six other patients also treated with BYL719. Acquired bi-allelic loss of *PTEN* was found in one of these patients, whereas in two others *PIK3CA* mutations present in the primary tumour were no longer detected at the time of progression. To characterize our findings functionally, we examined the effects of *PTEN* knockdown in several preclinical models (both in cell lines intrinsically sensitive to BYL719 and in *PTEN*-null xenografts derived from our index patient), which we found resulted in resistance to BYL719, whereas simultaneous PI(3)K p110 β blockade reverted this resistance phenotype. We conclude that parallel genetic evolution of separate metastatic sites with different *PTEN* genomic alterations leads to a convergent *PTEN*-null phenotype resistant to PI(3)K α inhibition.

We are currently engaged in testing the antitumour activity of a novel PI(3)K α inhibitor, BYL719, in patients with tumours harbouring activating PI(3)K α mutations³. The PI(3)K pathway is essential for cell growth, proliferation, survival, and metabolism^{4,5}. The PI(3)K family of enzymes is divided into three main classes (I to III), with class I being the most often implicated in human cancer⁶. Class IA PI(3)K is a heterodimer composed of a catalytic subunit (p110 α , β or δ) and a regulatory subunit^{7,8}. *PIK3CA*, the gene encoding p110 α , is mutated in up to 40% of oestrogen receptor (ER) and/or HER2 positive breast tumours^{9,10}. In our ongoing phase I clinical study of BYL719, we have observed clinical responses in breast, head and neck and other tumours³, providing proof

of principle that PI(3)K α targeting is active against tumours harbouring *PIK3CA* mutation.

We present the case of a 60-year-old breast cancer patient (index patient) diagnosed with invasive ductal carcinoma who underwent surgery followed by adjuvant treatment with chemotherapy and the aromatase inhibitor exemestane. Four years later, the patient developed bone metastases and started therapy with the ER antagonist fulvestrant, achieving stable disease. After 18 months on therapy, her disease progressed in the liver, bone and lymph nodes. The archival tissue of the primary tumour was subjected to PCR-based genetic analysis¹¹ and a hot spot mutation in *PIK3CA* (E542K) was detected. This finding led to the patient's enrolment in a phase I clinical trial testing the tolerability and antitumour activity of BYL719 (NCT01219699). The patient rapidly achieved a confirmed partial response according to the RECIST 1.0 criteria¹² that lasted 9.5 months (Table 1 and Extended Data Fig. 1). At that point, while the tumour remained stable in multiple sites including a peri-aortic lymph node location, progression occurred in the lungs (Fig. 1) and consequently

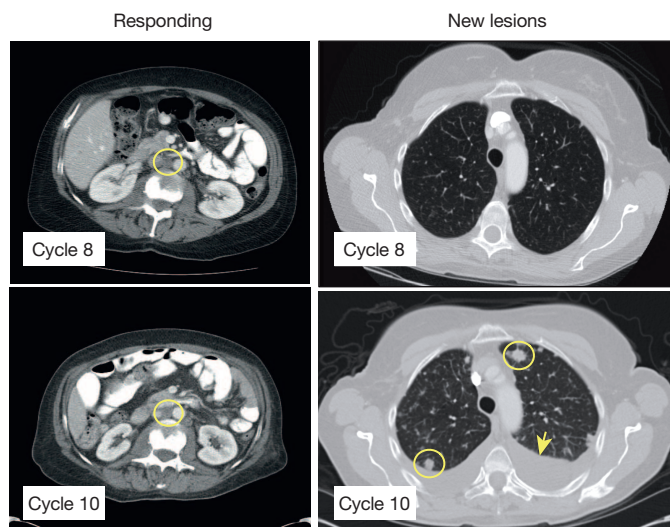


Figure 1 | Clinical response of index patient treated with BYL719. CT scans showing stable (responding) peri-aortic lymph node metastasis (yellow circles, left column) and the appearance of new lung metastatic lesions (yellow circles, right column) after the completion of the tenth cycle of BYL719 therapy. Arrow, pleural effusion.

¹Massachusetts General Hospital Cancer Center, 55 Fruit Street, Boston, Massachusetts 02114, USA. ²Human Oncology and Pathogenesis Program (HOPP), Memorial Sloan Kettering Cancer Center, 1275 York Avenue, Box 20, New York, New York 10065, USA. ³Department of Genetics, Washington University School of Medicine, 4566 Scott Avenue, St Louis, Missouri 63110, USA. ⁴Siteman Cancer Center, Washington University School of Medicine, 660 South Euclid Avenue, St Louis, Missouri 63110, USA. ⁵The Genome Institute, Washington University School of Medicine, 4444 Forest Park Avenue, St Louis, Missouri 63108, USA. ⁶Department of Medicine, Washington University School of Medicine, 660 South Euclid Avenue, St Louis, Missouri 63110, USA. ⁷Department of Pathology, Memorial Sloan Kettering Cancer Center, 1275 York Avenue, Box 20, New York, New York 10065, USA. ⁸Cancer Biology and Genetics Program, Memorial Sloan Kettering Cancer Center, 1275 York Avenue, Box 20, New York, New York 10065, USA. ⁹Division of Genitourinary Oncology, Department of Medicine, Memorial Sloan Kettering Cancer Center, 1275 York Avenue, Box 20, New York, New York 10065, USA. ¹⁰Howard Hughes Medical Institute, Cancer Biology and Genetics Program, Memorial Sloan Kettering Cancer Center, 1275 York Avenue, Box 20, New York, New York 10065, USA. ¹¹Novartis Pharma AG, Forum 1, Novartis Campus, CH-4056 Basel, Switzerland. ¹²Oncology Translational Medicine, Novartis Institutes for BioMedical Research, Cambridge, Massachusetts 02139, USA. ¹³Breast Medicine Service, Department of Medicine, Memorial Sloan Kettering Cancer Center, 1275 York Avenue, Box 20, New York, New York 10065, USA.

*These authors contributed equally to this work.

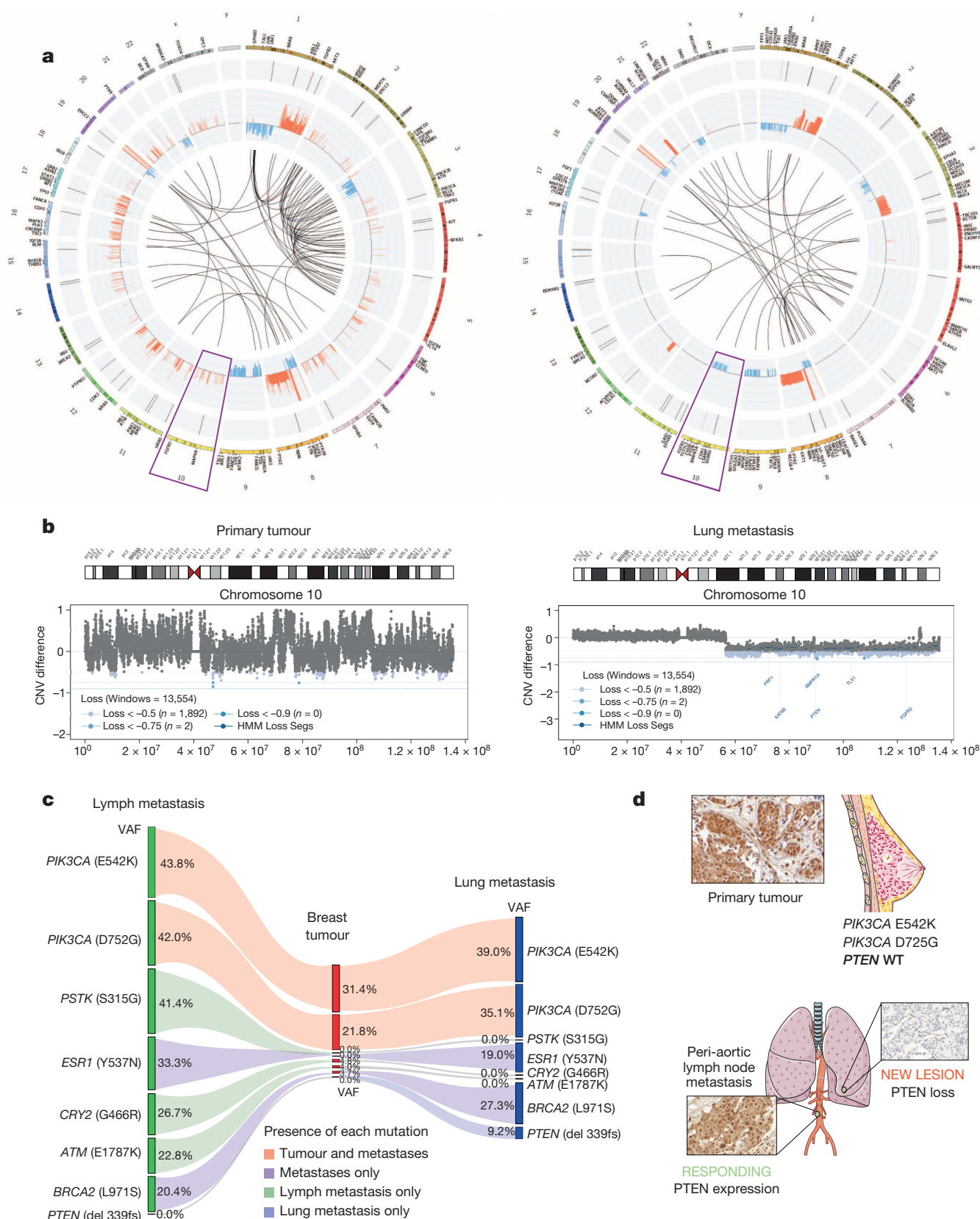


Figure 2 | Loss of PTEN upon BYL719 resistance. **a**, Circos plots from WGS analysis of primary tumour (before BYL719 treatment) and a lung metastasis appearing after the tenth cycle of BYL719 therapy. **b**, Copy number variation of chromosome 10. **c**, WES of the peri-aortic lymph node showing durable stable disease during BYL719 therapy compared to both primary tumour and the progressing lung lesion. The diagram shows the variant allele

fractions (VAF) of the listed gene mutations in the three lesions. The estimated tumour purities are 44% for the breast primary tumour, 50% for the lung metastasis, and 59% for the lymph node metastasis. **d**, PTEN IHC of primary tumour, peri-aortic lymph node, and lung metastasis. Images were taken from Servier Medical Art (licensed under a Creative Commons Attribution 3.0 Unported License).

Table 1 | Patient information

Patient	Site	PIK3CA baseline	Dose (mg)	Response (RECIST)	DOT (days)	PTEN Post-T	PIK3CA Post-T
Index	Breast	E542K	400	PR (−52.4%)	285	Loss	E542
1	Breast	H1047R	400	SD*	181	Unch.	H1047R
2	Breast	H1047R	400	SD (−26.3%)	424	Loss	H1047R
3	Breast	H1047R	400	SD (−28.5%)	179	Unch.	WT
4	Breast	H1047L	400	SD (−11.3%)	504	Unch.	H1047L
5	Breast	H1047R	400	SD (−24.9%)	110	Unch.	H1047R
6	Salivary	E545K	400	SD (−17%)	112	Unch.	WT

RECIST, Response Evaluation Criteria In Solid Tumors; DOT, duration of treatment; PR, partial response; SD, stable disease; Unch., unchanged; WT, wild-type; Amp., amplified; Post-T, post-treatment.

*With interval decrease in left breast and left chest wall skin thickening.

therapy with BYL719 was discontinued. The clinical status of the patient deteriorated rapidly and she died two months after termination of the BYL719 treatment. A rapid autopsy was performed three hours after death and a total of 14 metastases with tumour cells present were identified and collected for sequencing (Extended Data Table 1).

In order to proceed systematically to identify possible genetic determinants of acquired resistance to PI(3)K α inhibition, we took a three-step approach. First, we examined both the primary tumour (before BYL719 treatment) and the new lung metastasis by whole genome sequencing (WGS). Although both samples shared many somatic genetic aberrations (Fig. 2a and Extended Data Fig. 2), *PTEN* copy number loss was detected only in the lung metastasis (Fig. 2b). Second, we analysed by whole exome sequencing (WES) the primary tumour, lung metastasis, and the peri-aortic lesion that remained stable (responding) at the time of progression to BYL719 therapy (Fig. 2c). This analysis revealed that both peri-aortic and lung lesions harboured mutations in *PIK3CA*, *ESR1* and *BRCA2*, and single copy loss of *PTEN*. Importantly, in addition to the *PTEN* copy number loss, we identified a *PTEN* del339FS

(frameshift) mutation only in the lung metastasis (Fig. 2c). By immunohistochemistry (IHC), we observed that *PTEN* protein expression was lost in the lung metastasis but was present in both the primary tumour and peri-aortic lesion (Fig. 2d).

Third, to confirm and expand our findings, we sequenced the primary tumour and all the metastatic lesions to >500-fold coverage using a custom targeted deep-sequencing assay, IMPACT^{13,14} (Methods). A number of mutations were shared by the primary tumour and the metastatic sites, whereas others were observed only in all or in selected metastatic lesions (Fig. 3a and Supplementary Table). We confirmed that the *PIK3CA* E542K mutation in the primary tumour was conserved in the metastatic samples and detected the presence of another *PIK3CA* mutation (D725G). Moreover, we found increased copy number of *FGFR1* and *E14EBP1* in all tumour samples, consistent with the relatively frequent 8p11-12 amplification described in breast cancer^{15,16}. *ESR1* Y537N and *BRCA2* L971S alterations were present in all the metastatic lesions but not in the primary tumour. We speculate that the *ESR1* Y537N mutation, reported to promote ligand-independent ER activation¹³, was selected upon anti-oestrogen therapy received by the patient before BYL719 treatment.

Central to our work, all metastatic lesions appeared to harbour a single copy loss of *PTEN* (Extended Data Fig. 3). Furthermore, we found that 10 of the 14 metastatic lesions harboured additional genomic alterations within *PTEN*. The spectrum of *PTEN* alterations was heterogeneous across the 10 samples and included a splice site mutation at K342, a frameshift indel at P339 (confirming the WES result), and 4 different exon-level deletions (Fig. 3a and Extended Data Fig. 4). All 10 specimens with either secondary *PTEN* mutations or copy number loss were confirmed negative for *PTEN* staining by IHC, whereas the four specimens that retained a *PTEN* wild-type allele were positive for *PTEN* protein expression (Extended Data Fig. 5). In addition, for those lesions

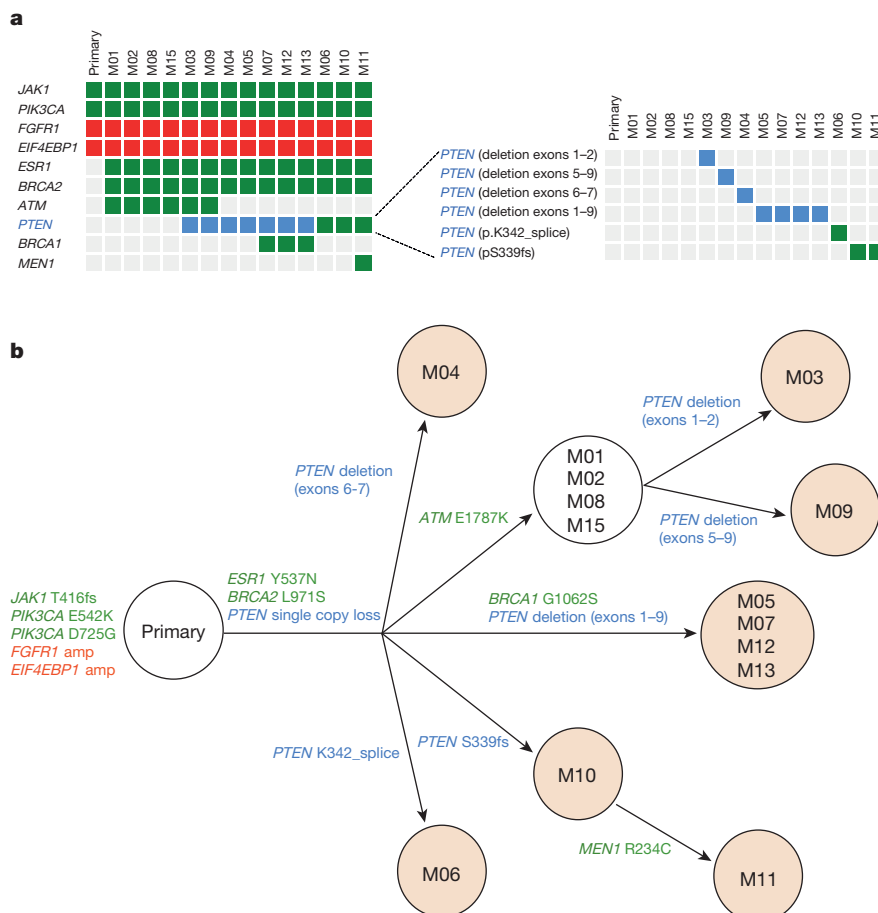


Figure 3 | Loss of *PTEN* by different genetic alterations. **a**, Heatmap of the non-silent genetic alterations across the primary tumour and the 14 metastases (M) collected during the autopsy of the index patient. Gene mutations are depicted in green, gene amplifications in red, and gene copy number loss in blue. **b**, Dendrogram showing the proposed phylogenetic evolution of the metastases in the index patient. Shaded circles represent metastatic lesions with bi-allelic loss of *PTEN* and lack of *PTEN* expression by IHC.

that were visualized by CT (computerized tomography) scan, there was a tight correlation between progression of disease and loss of PTEN expression. The peri-aortic lesion (M02) that was responding at the time of disease progression still contained one *PTEN* wild-type allele and protein expression. Conversely, the lung lesions (M04, M06, M09 and M11) and liver lesion (M12) with documented progression to therapy had bi-allelic *PTEN* alteration and lack of expression.

In an effort to integrate the genomic data from our patient, we constructed a dendrogram mapping the phylogenetic evolution of the disease. Our findings suggest that all the lesions were derived from the *PTEN* wild-type primary tumour, and that there was a progressive and parallel loss of PTEN under BYL719 selective pressure (Fig. 3b). Of note, the two-month duration between progression to BYL719 and autopsy needs to be considered as well.

In order to expand our observations, we analysed paired samples (pre-treatment and at progression) from six additional patients enrolled in the same study at our institution (Table 1). Targeted sequencing identified homozygous loss of *PTEN* in a post-treatment sample of a breast cancer patient who developed resistance to BYL719 after initially experiencing a durable response to therapy (Table 1). We also confirmed lack of PTEN expression by IHC in the post-treatment sample (Extended Data Fig. 6). We found no detectable *PIK3CA* mutations in the post-treatment samples of two patients (Table 1). Given that the presence of

PIK3CA mutations drove sensitivity to BYL719 in our cell line screens¹⁷, positive selection of clones bearing wild-type alleles of *PIK3CA* may explain the emergence of resistance to BYL719 in these two additional cases. These results may be an indication that in some cases *PIK3CA* mutations are not early founder/truncal events but branched subclonal drivers that are cleared from the tumours under the selective pressure of PI(3)K α inhibition. In any case, the fact that loss of PTEN expression and emergence of *PIK3CA* wild-type clones are mutually exclusive in our patient samples indicates that both events may be important in opposing the therapeutic efficacy of BYL719.

No other alterations with an obvious connection to BYL719 resistance were found in the responding cases, with the exception of a mutation (E1490*) and an in-frame deletion in *MAP3K1* in one of the three patients for whom neither PTEN nor *PIK3CA* status changed during BYL719 treatment. Further characterization is needed to determine whether these mutations lead to increased MEK and ERK signaling and limit the effects of PI(3)K inhibition.

PTEN encodes for a phosphatase that regulates the activity of PI(3)K by limiting the accumulation of phosphatidylinositol-3,4,5-trisphosphate (PtdIns(3,4,5)P₃ or PIP₃), a required mediator to initiate the PI(3)K/AKT/mTOR signalling cascade¹⁸. In the absence of PTEN, cancer cells become dependent mostly on the activity of the p110 β isoform of PI(3)K (PI(3)K β) to propagate signalling through downstream pathway

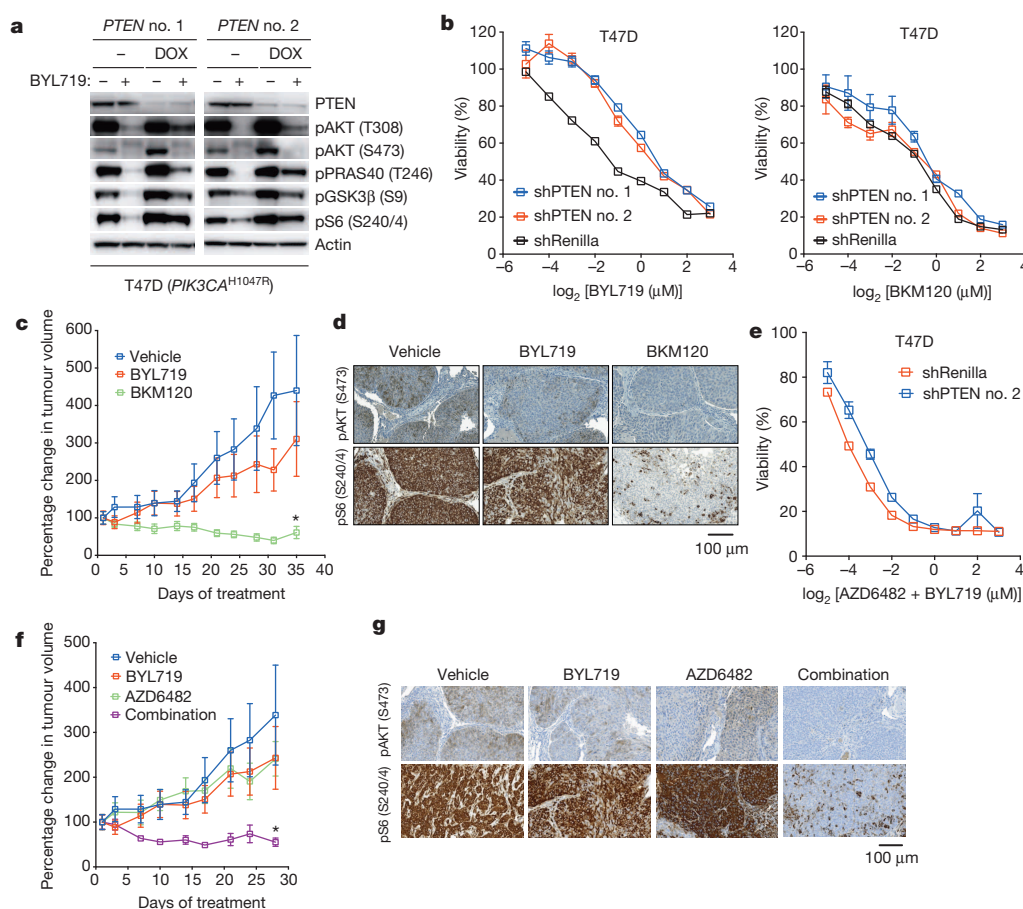


Figure 4 | Loss of PTEN expression and sensitivity to PI(3)K α and PI(3)K β blockade. **a**, Western blot showing PTEN knockdown by two independent shRNAs and its effects on the PI(3)K/AKT/mTOR pathway. **b**, Cell viability assay in T47D cells with inducible PTEN knockdown (shPTEN no. 1 and no. 2) or PTEN expressing controls (shRenilla) treated with increasing concentrations of BYL719 or BKM120. Error bars, s.e.m. **c**, Antitumour activity of either BYL719 (25 mg kg⁻¹ daily) or BKM120 (25 mg kg⁻¹ daily) in PDXs subcutaneously grown in nude mice ($n = 6$ (vehicle) and $n = 8$ (treatments)). Error bars, s.e.m. **d**, Representative immunostaining for phosphorylated AKT (pAKT) and phosphorylated S6 (pS6) in PDXs treated as shown. Tumours were

collected at the end of the experiment of **c**, 2 h after the last dosage. Scale bar, 100 μ m. **e**, Cell viability assay in T47D cells with PTEN expression or PTEN knockdown treated with increasing concentrations of the combination of BYL719 and AZD6482. Error bars, s.e.m. **f**, Antitumour activity of either BYL719 (25 mg kg⁻¹ daily) or the combination of BYL719 and AZD6482 (25 mg kg⁻¹ daily) in PDXs subcutaneously grown in nude mice ($n = 6$ (vehicle) and $n = 8$ (treatments)). Error bars, s.e.m. **g**, Representative immunostaining for pAKT (S473) and pS6 (S240/244) in PDXs treated as shown. Tumours were collected at the end of the experiment of **f**, 2 h after the last dosage. Scale bar, 100 μ m. * $P < 0.05$.

effectors^{19,20}. Therefore, we hypothesized that progressive decrease or loss of PTEN expression in the presence of PI(3)K α inhibition might restore PI(3)K/AKT signalling through PI(3)K β activity. To test our hypothesis, we established cell lines expressing either doxycycline-inducible or constitutive short hairpin (sh)RNA against PTEN messenger RNA using T47D and MCF7 cells, known to be intrinsically sensitive to BYL719²¹. As expected, induction of PTEN downregulation led to activation of AKT and the downstream effectors PRAS40, GSK3 β and S6 in both T47D (Fig. 4a) and MCF7 (data not shown) cells under basal conditions. PTEN downregulation markedly limited the effects of BYL719, both at the signalling and cell viability level. On the other hand, PTEN knockdown did not result in resistance to the pan-PI(3)K inhibitor BKM120, which blocks all the PI(3)K p110 isoforms (Fig. 4b and Extended Data Fig. 7a). Similar effects were observed in another BYL719-sensitive cell line (MDA-MB-453) with constitutive PTEN knockdown (Extended Data Fig. 7b).

From our patient's non-responding PTEN-null lung metastatic lesion, we were able to establish xenografts in nude mice. Consistent with the *in vitro* data, this patient-derived xenograft (PDX) was resistant to BYL719 treatment but sensitive to BKM120 (Fig. 4c). The degree of inhibition of phospho-AKT and phospho-S6 was also higher with BKM120 (Fig. 4d and Extended Data Fig. 7c and d). These results were complemented by the combination of BYL719 and the PI(3)K β inhibitor AZD6482. Upon PTEN knockdown, only combined PI(3)K α and β blockade was capable of reverting the resistant phenotype (Fig. 4e and Extended Data Fig. 8a). Similarly, the BYL719-resistant PDX was insensitive to AZD6482 alone but responded to the combination of both compounds (Fig. 4f). Profound inhibition of AKT and S6 phosphorylation was achieved only upon treatment with BYL719 in combination with AZD6482 (Fig. 4g and Extended Data Fig. 8b and c). Taken together, these data indicate that inhibition of the PI(3)K β isoform is required to achieve antitumour activity in cells/tumours that lost PTEN expression and become resistant to BYL719.

We have reported a case of parallel genetic evolution under selective therapeutic pressure leading to a progressive loss of PTEN expression and consequent gain of dependency on the PI(3)K β isoform. Parallel evolution under selective pressure has been described in conditions where treatments are highly efficacious, such as in HIV²². Our case highlights that this tumour, despite its heterogeneity, was dependent on PI(3)K signalling, probably as a result of the presence of the same activating *PIK3CA* mutation in all the tumour sites. Upon continued suppression of PI(3)K α , diverse genomic alterations emerged, leading to PTEN loss as an alternative mechanism of PI(3)K activation. Moreover, our study emphasizes the importance of tumour interrogation upon progression to therapy and the dynamic nature of tumour genomes under selective therapeutic pressure.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 29 April; accepted 7 October 2014.

Published online 17 November 2014.

- Gerlinger, M. *et al.* Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N. Engl. J. Med.* **366**, 883–892 (2012).
- Swanton, C. Intratumor heterogeneity: evolution through space and time. *Cancer Res.* **72**, 4875–4882 (2012).
- Juric, D. BYL719, a next generation PI3K α specific inhibitor: preliminary safety, PK, and efficacy results from the first-in-human study. *Cancer Res.* **72**, Abstr. CT-01 (2012).

- Engelman, J. A., Luo, J. & Cantley, L. C. The evolution of phosphatidylinositol 3-kinases as regulators of growth and metabolism. *Nature Rev. Genet.* **7**, 606–619 (2006).
- Cantley, L. C. The phosphoinositide 3-kinase pathway. *Science* **296**, 1655–1657 (2002).
- Engelman, J. A. Targeting PI3K signalling in cancer: opportunities, challenges and limitations. *Nature Rev. Cancer* **9**, 550–562 (2009).
- Brachmann, S. M., Ueki, K., Engelman, J. A., Kahn, R. C. & Cantley, L. C. Phosphoinositide 3-kinase catalytic subunit deletion and regulatory subunit deletion have opposite effects on insulin sensitivity in mice. *Mol. Cell. Biol.* **25**, 1596–1607 (2005).
- Zhao, L. & Vogt, P. K. Helical domain and kinase domain mutations in p110 α of phosphatidylinositol 3-kinase induce gain of function by different mechanisms. *Proc. Natl Acad. Sci. USA* **105**, 2652–2657 (2008).
- Miller, T. W., Rexer, B. N., Garrett, J. T. & Arteaga, C. L. Mutations in the phosphatidylinositol 3-kinase pathway: role in tumor progression and therapeutic implications in breast cancer. *Breast Cancer Res.* **13**, 224–235 (2011).
- Samuels, Y. *et al.* High frequency of mutations of the *PIK3CA* gene in human cancers. *Science* **304**, 554 (2004).
- Sequist, L. V. *et al.* Implementing multiplexed genotyping of non-small-cell lung cancers into routine clinical practice. *Ann. Oncol.* **22**, 2616–2624 (2011).
- Therasse, P. *et al.* New guidelines to evaluate the response to treatment in solid tumors. European Organization for Research and Treatment of Cancer, National Cancer Institute of the United States, National Cancer Institute of Canada. *J. Natl. Cancer Inst.* **92**, 205–216 (2000).
- Toy, W. *et al.* ESR1 ligand-binding domain mutations in hormone-resistant breast cancer. *Nature Genet.* **45**, 1439–1445 (2013).
- Won, H. H., Scott, S. N., Brannon, A. R., Shah, R. H. & Berger, M. F. Detecting somatic genetic alterations in tumor specimens by exon capture and massively parallel sequencing. *J. Visual. Expts* e50710 (2013).
- Yang, Z. Q., Liu, G., Bollig-Fischer, A., Giroux, C. N. & Ethier, S. P. Transforming properties of 8p11-12 amplified genes in human breast cancer. *Cancer Res.* **70**, 8487–8497 (2010).
- Karlsson, E. *et al.* High-resolution genomic analysis of the 11q13 amplicon in breast cancers identifies synergy with 8p12 amplification, involving the mTOR targets S6K2 and 4EBP1. *Genes Chromosom. Cancer* **50**, 775–787 (2011).
- Fritsch, C. *et al.* Characterization of the novel and specific PI3K α inhibitor NVP-BYL719 and development of the patient stratification strategy for clinical trials. *Mol. Cancer Ther.* **13**, 1117–1129 (2014).
- Stambolic, V. *et al.* Negative regulation of PKB/Akt-dependent cell survival by the tumor suppressor PTEN. *Cell* **95**, 29–39 (1998).
- Jia, S. *et al.* Essential roles of PI(3)K-p110 β in cell growth, metabolism and tumorigenesis. *Nature* **454**, 776–779 (2008).
- Edgar, K. A. *et al.* Isoform-specific phosphoinositide 3-kinase inhibitors exert distinct effects in solid tumors. *Cancer Res.* **70**, 1164–1172 (2010).
- Elkabets, M. *et al.* mTORC1 inhibition is required for sensitivity to PI3K p110 α inhibitors in PIK3CA-mutant breast cancer. *Sci. Transl. Med.* **5**, 196ra199 (2013).
- Lemey, P. *et al.* Molecular footprint of drug-selective pressure in a human immunodeficiency virus transmission chain. *J. Virol.* **79**, 11981–11989 (2005).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank members of the MSKCC Diagnostic Molecular Pathology Laboratory and the MSK Maria-Josée and Henry Kravis Center for Molecular Oncology for assistance with sequencing. We thank M. Asher and U. Bhanot from the MSKCC Pathology Core for assistance with tissue staining. This work was funded by a “Stand Up to Cancer” Dream Team Translational Research Grant, a Program of the Entertainment Industry Foundation (SU2C-AACR-DT0209), the Breast Cancer Research Foundation, the Geoffrey Beene Cancer Research Center, the Starr Cancer Consortium and an MMHCC grant (CA105388). D.J. is also funded by a National Institutes of Health Training Grant (T32 CA-71345-15) and by philanthropic support from Stephen and Kathleen Chubb.

Author Contributions D.J., P.C., M.F.B., J.B. and M.S. conceived the project, designed and analysed the experiments, and wrote the manuscript. M.G., O.L.G., B.J.A., A.R. and E.R.M. performed and analysed the WGS and WES data. T.H., M.M.-K., D.S., S.I., A.T., L.E., C.Q., M.P., A.D., R.B. and A.H. collected and analysed patients' samples. P.C., H.E., S.H.E. and S.W.L. performed and supervised the laboratory experiments. H.H.W., G.I., R.H.S., D.B.S. and M.F.B. performed and supervised the IMPACT sequencing and analysis.

Author Information DNA sequences have been deposited in the European Genome-phenome Archive with accession number EGAS00001000991. Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.S. (scaltrim@mskcc.org), J.B. (baselgaj@mskcc.org) or M.F.B. (bergerm1@mskcc.org).

METHODS

PIK3CA mutant cell lines MCF7 (E545K) and T47D (H1047R) (ATCC) were transduced with the retroviral TRMPV vector. Doxycycline (Sigma) was used to temporally activate the expression of a microRNA 30-embedded shRNA targeting *Renilla luciferase* (control) or *PTEN* mRNA. The hairpin sequences used were as follows.

Renilla luciferase. CTCGAGAAGGTATATTGCTGTTGACAGTGAGCGCAG GAATTATAATGCTTATCTATAGTGAAGCCACAGATGTATAGATAAGCA TTATAATTCCTATGCCTACTGCCTCGGAATTC

PTEN no. 1. CTCGAGAAGGTATATTGCTGTTGACAGTGAGCGACAGC TAAAGGTGAAGATATATAGTGAAGCCACAGATGTATATATCTTCACCT TTAGCTGGCTGCCTACTGCCTCGGAATTC

PTEN no. 2. CTCGAGAAGGTATATTGCTGTTGACAGTGAGCGCCAGAT GTTAGTGACAATGAATAGTGAAGCCACAGATGTATTCAATTGTCTACTAA CATCTGGTTGCCTACTGCCTCGGAATTC

Cell viability was assessed using the tetrazolium-based MTT assay after 6 days of treatment. All cell lines resulted negative for mycoplasma contamination. Western blotting was carried out using previously described methods²¹. All the *in vitro* experiments were performed in triplicate.

Patient-derived xenografts and IHC. Animals were maintained and treated in accordance with Institutional Guidelines of Memorial Sloan Kettering Cancer Center (Protocol number 12-10-019). Tumours were implanted subcutaneously in six-week-old female athymic NU/NU nude mice. Once the tumours reached a volume of ~200 mm³, it was expanded in multiple mice which were then randomized to the following treatments: BYL719, BKM120 (a pan-class I PI(3)K inhibitor), or AZD6482 (a PI(3)K β inhibitor), each administered orally at 25 mg kg⁻¹ once a day. After treatment, mice were euthanized and tumours were harvested and procured for IHC analysis. IHC was performed on a Ventana Discovery XT processor platform using standard protocols and the following antibodies: pAKT (S473) (D9E), Cell Signaling Technology, #4060, dilution 1:70. pS6 (S240/4) (D68F8)XP, Cell Signaling Technology, #5364, dilution 1:500. PTEN (138G6), Cell Signaling Technology, #9559, dilution 1:30.

All the *in vivo* experiments were run with at least $n = 8$ for each treatment arm. Two-way *t*-test was performed using GraphPad Prism (GraphPad Software). Error bars represent the s.e.m. * $P < 0.05$.

Whole genome and whole exome sequencing. For whole genome (WG) and whole exome (WE) sequencing, DNA was derived from the primary tumour, lung metastasis, and peri-aortic lymph node metastasis. DNA from the spleen was used as a normal control. WG libraries were produced as previously described²³ and sequenced using the Illumina HiSeq 2500 platform as paired-end 100 base pair reads, producing ~30-fold (primary tumour, spleen normal)-50-fold (lung metastasis) coverage for WG sequencing. By hybrid capture (Nimblegen version 3.0) of the lymph node and lung metastases, primary tumour and spleen normal, we generated ~100-fold coverage for WE sequencing.

All patients provided written informed consent for the genetic research studies performed in accordance with protocols approved by Dana Farber/Harvard Cancer

Center Institutional Review Board. Autopsy in the index patient was performed within the first three hours *post mortem*.

Targeted exome sequencing (IMPACT). DNA derived from the primary tumour, 14 metastases, and matched normal spleen tissue was further subjected to deep-coverage targeted sequencing of key cancer-associated genes. Our assay, termed IMPACT (Integrated Mutation Profiling of Actionable Cancer Targets), involves hybridization of barcoded libraries to custom oligonucleotides (Nimblegen SeqCap) designed to capture all protein-coding exons and select introns of 279 commonly implicated oncogenes, tumour suppressor genes, and members of pathways deemed actionable by targeted therapies¹⁴. The captured pool was subsequently sequenced on an Illumina HiSeq 2500 as paired-end 75-base pair reads, producing 513-fold coverage per tumour. Sequence data were analysed to identify three classes of somatic alterations: single-nucleotide variants, small insertions/deletions (indels), and copy number alterations.

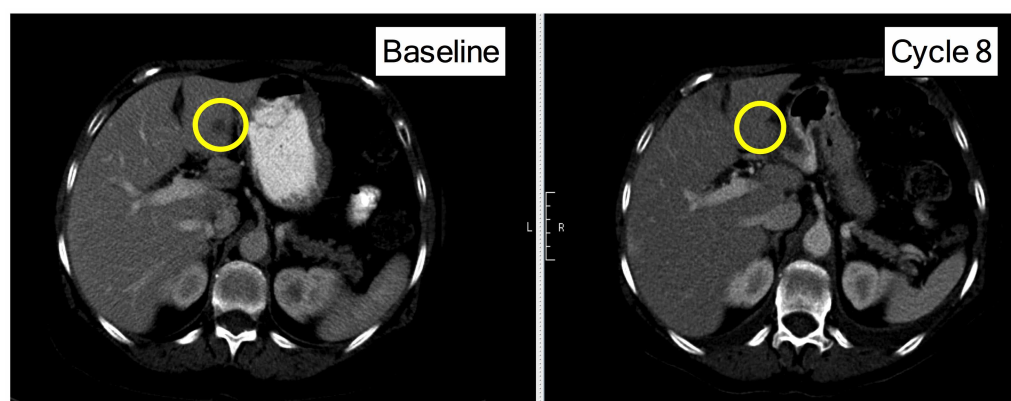
Barcoded sequence libraries were prepared using 250 ng genomic DNA (Kapa Biosystems) and combined in a single equimolar pool. Sequence data were demultiplexed using CASAVA, and reads were aligned to the reference human genome (hg19) using BWA and postprocessed using the Genome Analysis Toolkit (GATK) according to GATK best practices^{24,25}.

MuTect and GATK were used to call single-nucleotide variants and small indels, respectively²⁶. Exon-level copy number gains and losses were inferred from the ratio in Tumour:Normal sequence coverage for each target region, following a loss-normalization to adjust for the dependency of coverage on GC content²⁷.

Statistical analysis. Two-way *t*-tests were performed using GraphPad Prism (GraphPad Software). Error bars represent the s.e.m., *P* values are indicated as * $P < 0.05$. All cellular experiments were repeated at least three times. All the *in vivo* experiments were run with at least 6–8 tumours for each treatment arm. Sample size was chosen to detect a difference in means of 20% with a power of 90%. Animals were randomized in groups with similar average in tumour size. Investigators were blinded when assessing the outcome of the *in vivo* experiments.

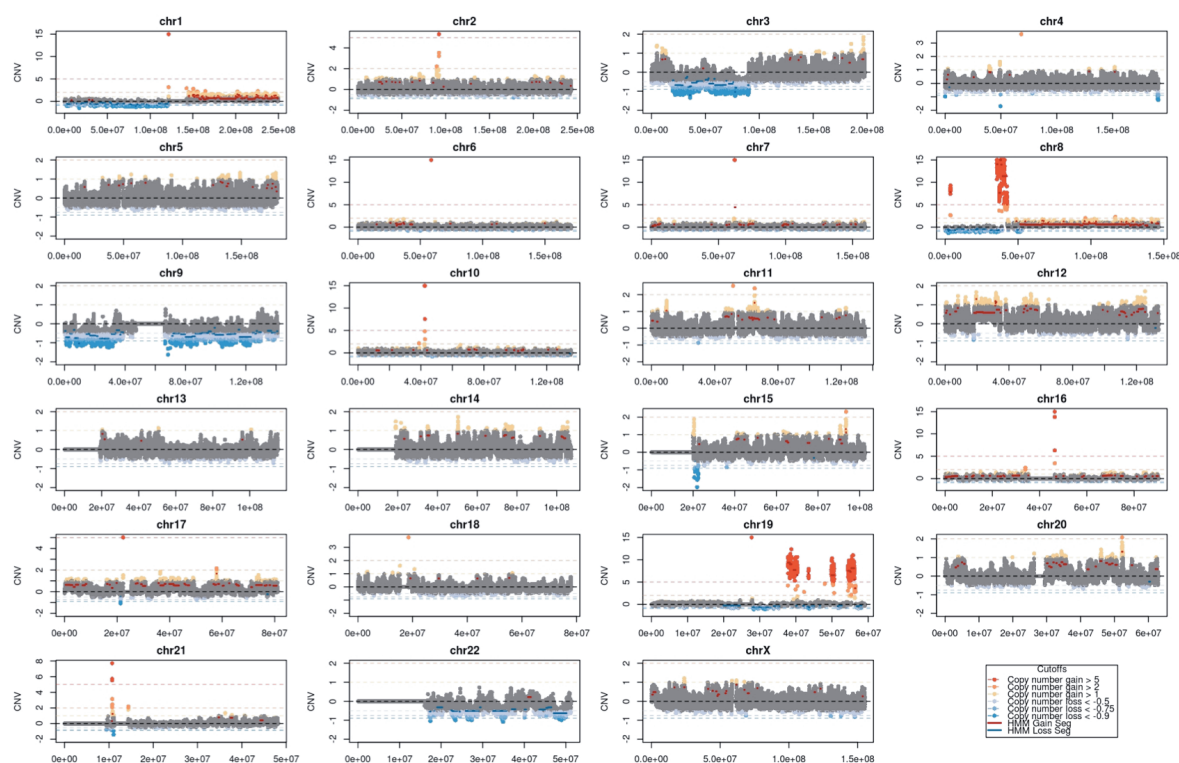
For the cell viability graphs, nonlinear regression was applied to the experimental data sets. Curves were compared using the extra-sum-of-squares *F* test using $\alpha = 0.05$. Hypothesis was rejected when nonlinear models were not nested within each other and was considered statistically significant.

23. Li, S. *et al.* Endocrine-therapy-resistant *ESR1* variants revealed by genomic characterization of breast-cancer-derived xenografts. *Cell Rep.* **4**, 1116–1130 (2013).
24. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
25. DePristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genet.* **43**, 491–498 (2011).
26. Cibulskis, K. *et al.* Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nature Biotechnol.* **31**, 213–219 (2013).
27. Wagle, N. *et al.* High-throughput detection of actionable genomic alterations in clinical tumor samples by targeted, massively parallel sequencing. *Cancer Discov.* **2**, 82–93 (2012).

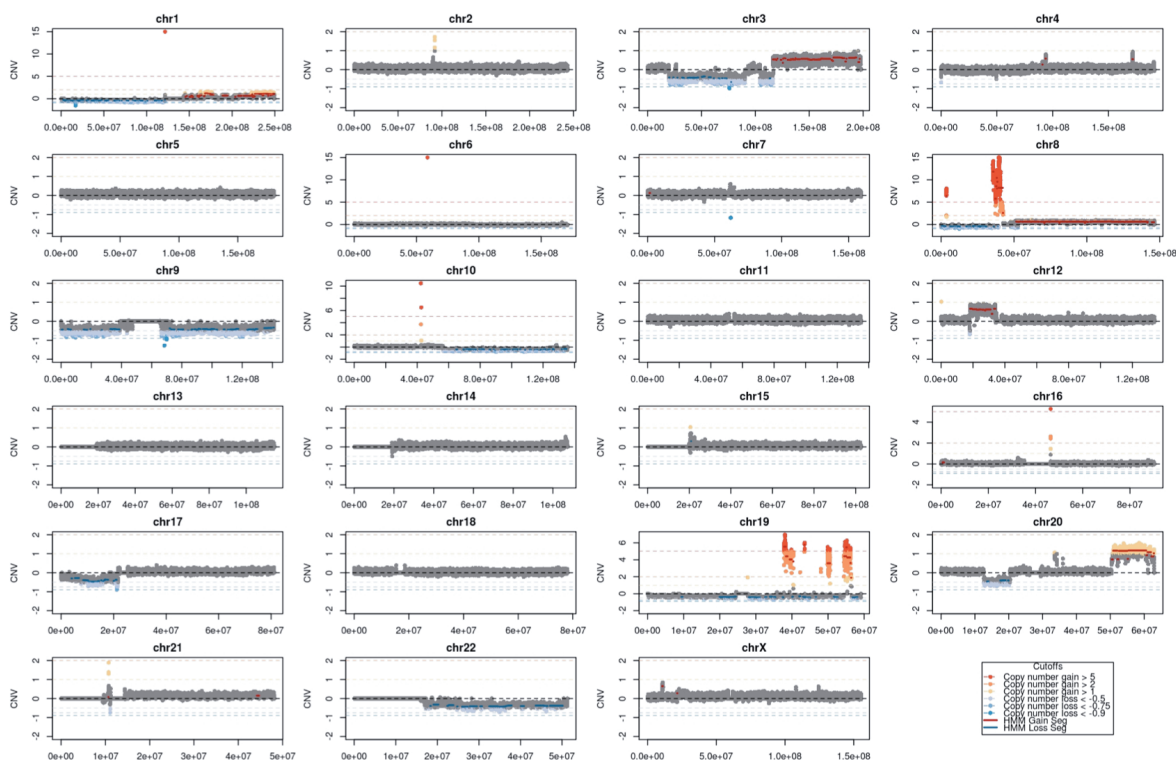


Extended Data Figure 1 | CT scan of index patient. CT scan showing a liver lesion (baseline) experiencing a partial response after 8 cycles (cycle 8) of BYL719.

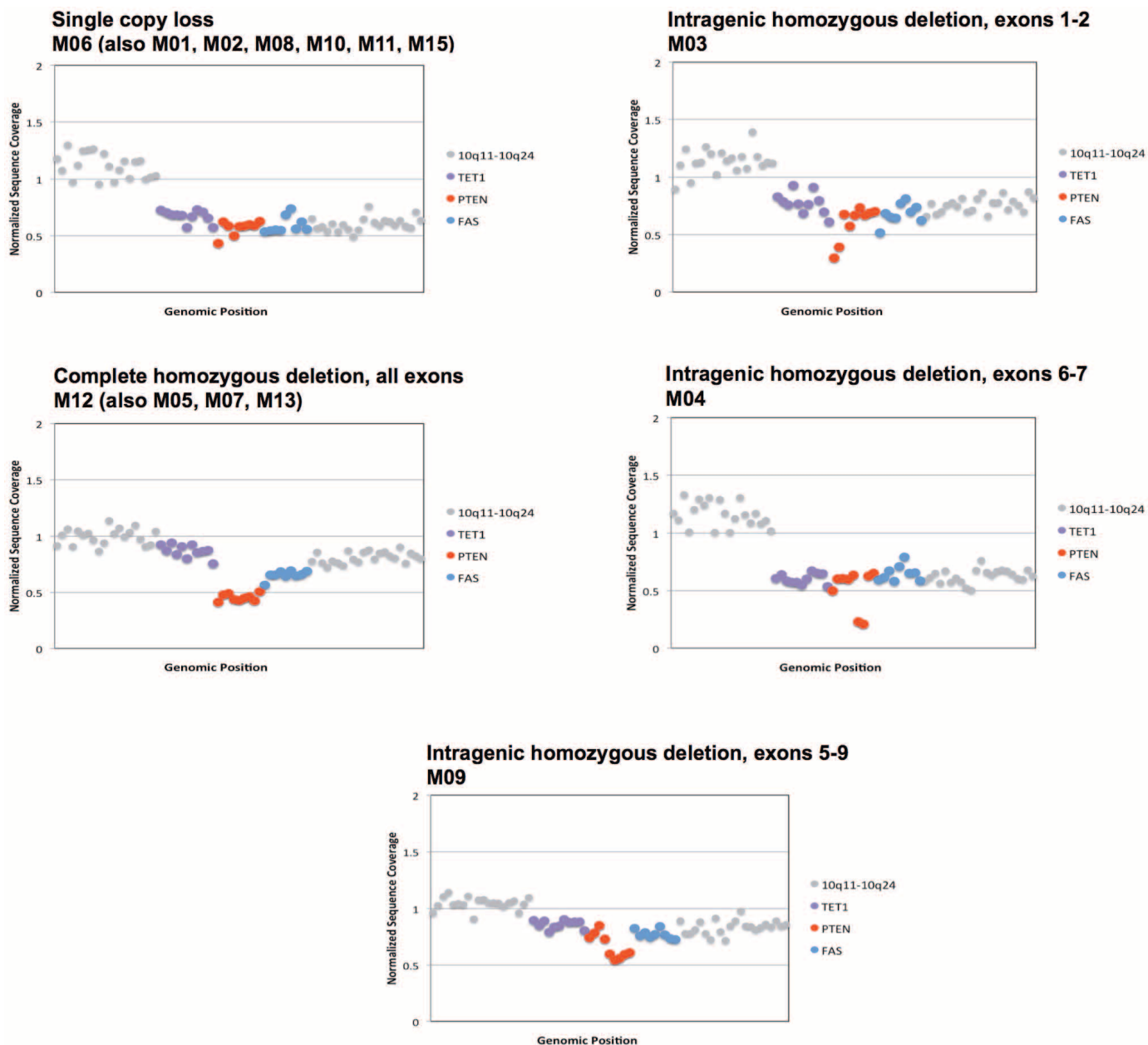
Primary Tumour



Lung metastasis

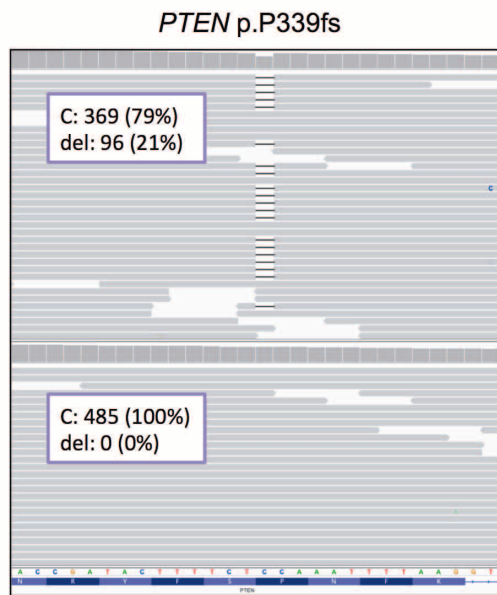


Extended Data Figure 2 | Gene copy number variation in both primary tumour and lung metastasis.

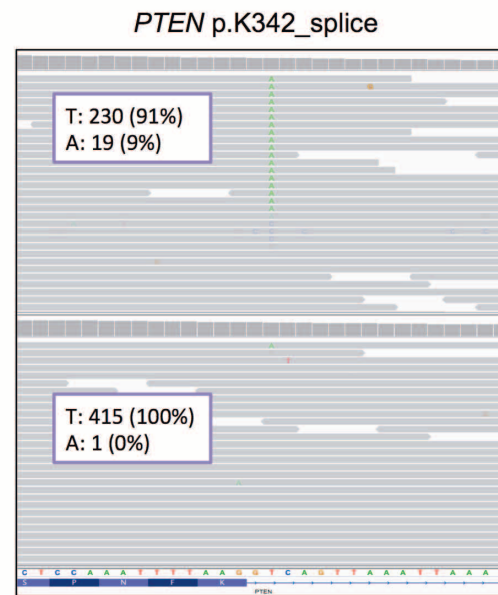


Extended Data Figure 3 | Representative exon-level copy number profiles for genes on chromosome 10 in all 14 metastases collected from the index patient. Exons in *PTEN* are shown in red.

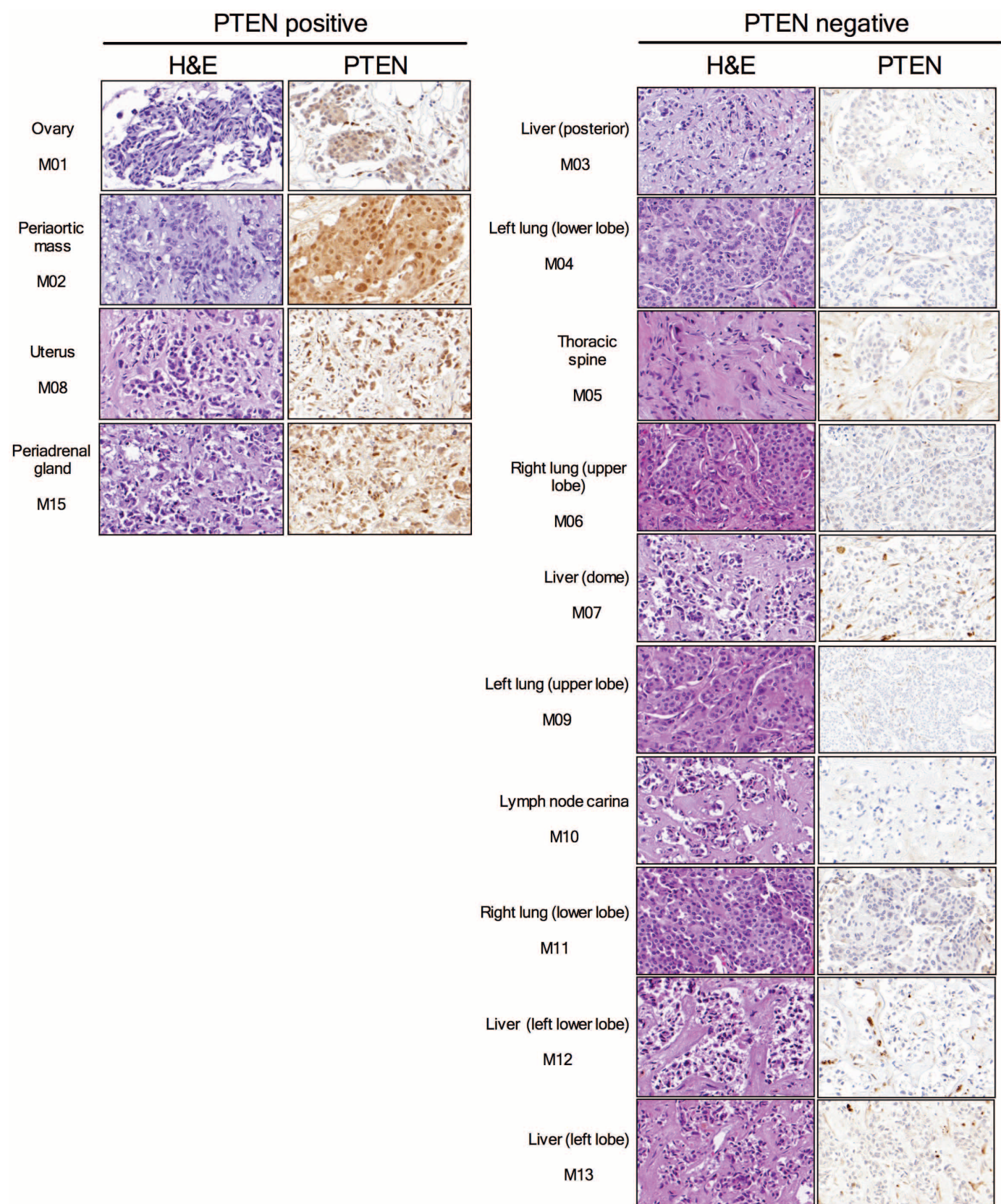
Tumour M10



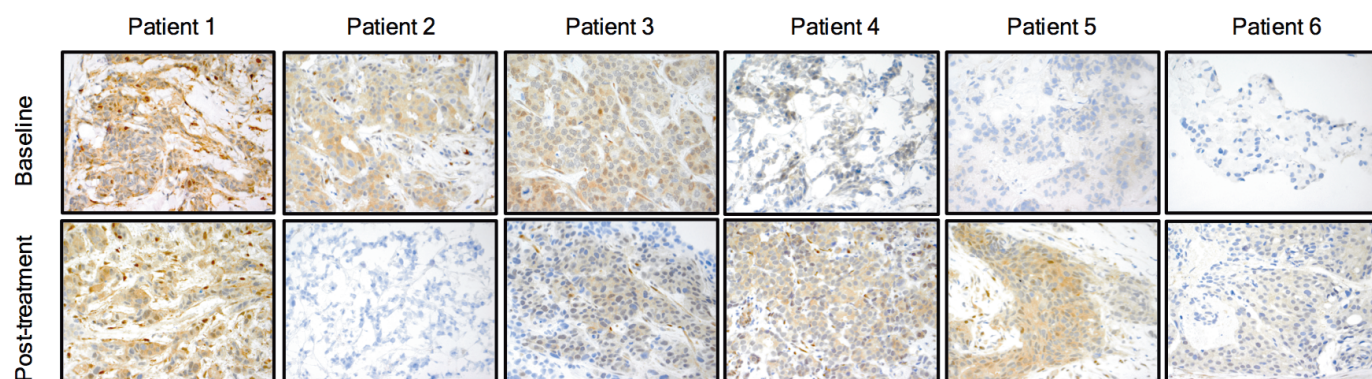
Tumour M06



Extended Data Figure 4 | Loss-of-function mutations in *PTEN* detected by IMPACT in metastases M06 and M10. Mutations were visualized by the Integrative Genomics Viewer (IGV).

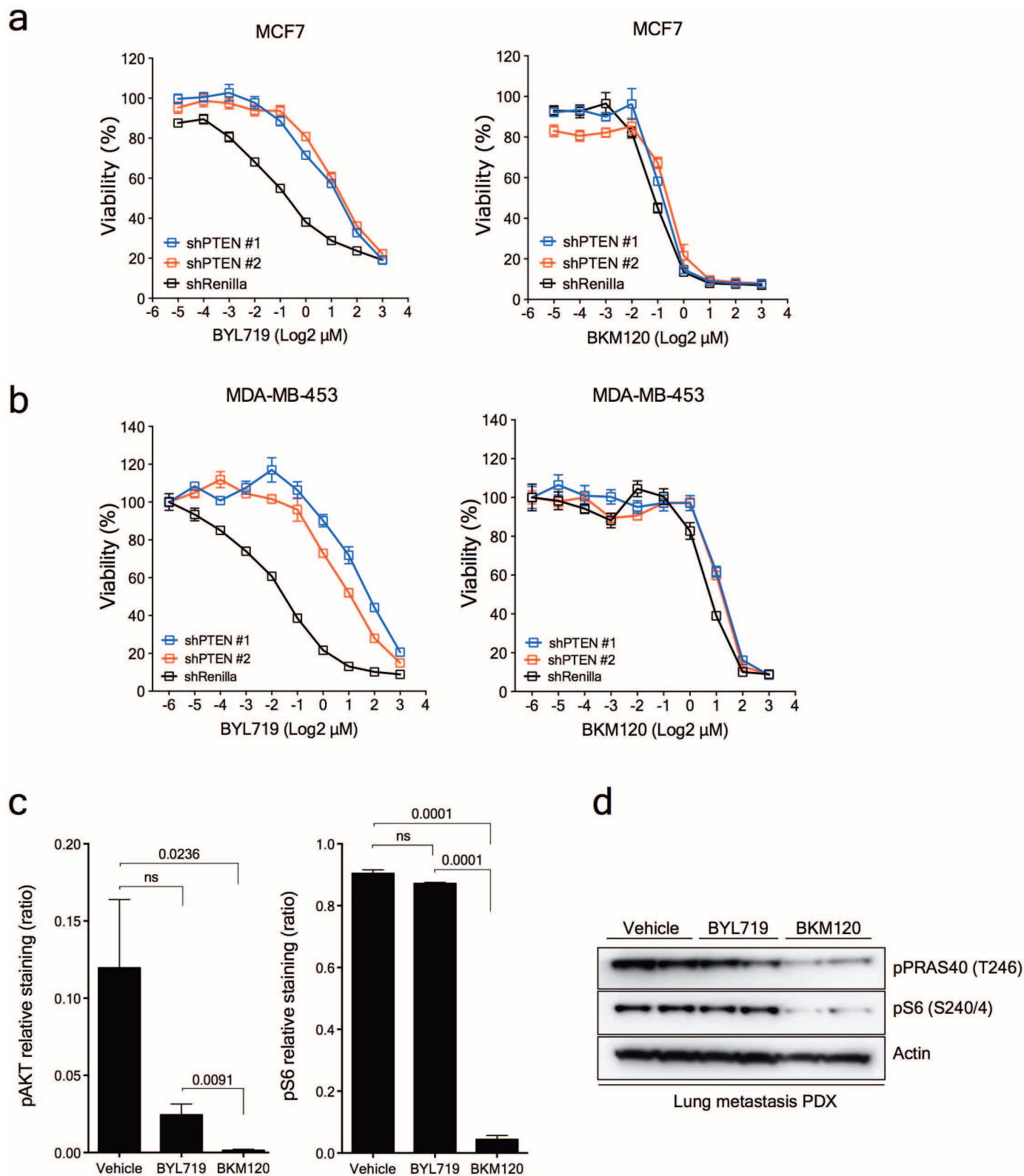


Extended Data Figure 5 | PTEN immunostaining of the 14 metastases collected during the autopsy. Haematoxylin and eosin (H&E) and PTEN expression detected by IHC in 14 metastases collected during the autopsy of the index patient. PTEN staining in PTEN negative samples is only present in stromal cells.



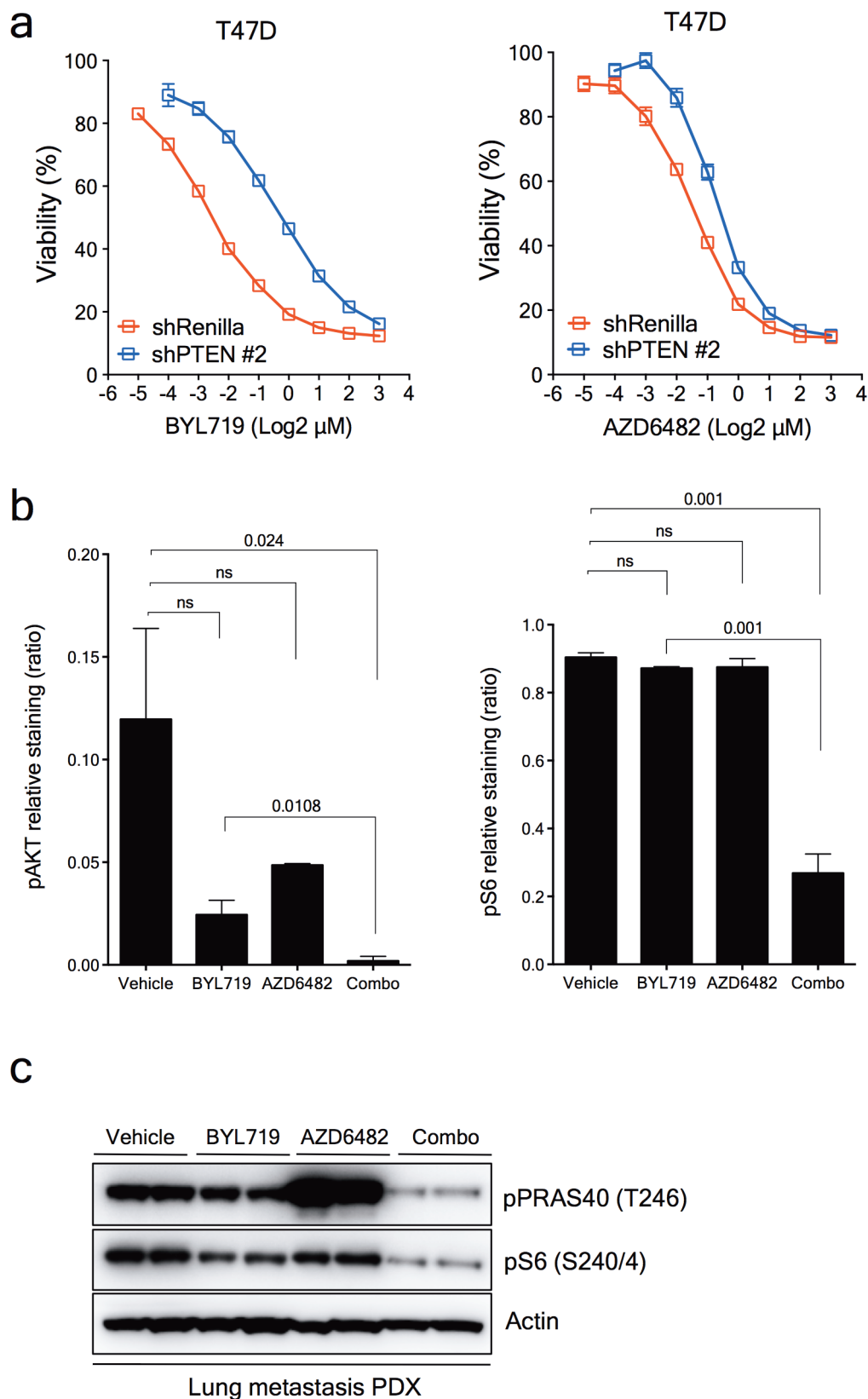
Extended Data Figure 6 | PTEN immunostaining in patients treated with BYL719. PTEN expression detected by IHC in paired samples from six additional patients treated with BYL719. Specimens before starting BYL719

therapy (baseline) and at time of disease progression (post-treatment) are compared.



Extended Data Figure 7 | Inducible loss of PTEN and sensitivity to BYL719 and BKM120. **a.** Cell viability assay in MCF7 cells with inducible *PTEN* knockdown treated with increasing concentrations of either BYL719 or BKM120. Error bars, s.e.m. **b.** Cell viability assay in MDA-MB-453 (MDA453) cells with constitutive *PTEN* knockdown treated with increasing

concentrations of either BYL719 or BKM120. Error bars, s.e.m. **c.** Quantification of pAKT (S473) and pS6 (S240/4) from Fig. 4d. Student's *t*-test was used and *P* values are indicated. **d.** Western blot from the PDXs treated as indicated.



Extended Data Figure 8 | Constitutive loss of PTEN and sensitivity to BYL719 and AZD6482. **a**, Cell viability assay in T47D cells with inducible *PTEN* knockdown (no. 2) treated with increasing concentrations of either BYL719 or AZD6482 in the presence of doxycycline 1 μg ml⁻¹. Error bars,

s.e.m. **b**, Quantification of pAKT (S473) and pS6 (S240/4) from Fig. 4g. Student's *t*-test was used and *P* values are indicated. Error bars, s.e.m. **c**, Western blot from the PDXs treated as indicated.

Extended Data Table 1 | Samples analysed from the index patient

Lesion	Location	Cellularity	CT scan	Response	WGS	WES	IMPACT
Primary	Breast	Unknown			Y	Y	Y
M01	Ovary	75%	N		N	N	Y
M02	Periaortic lymph node	50%	Y	Y	Y	Y	Y
M03	Liver (posterior)	65%	N		N	N	Y
M04	Left Lung (lower lobe)	75%	Y	N	N	N	Y
M05	Thoracic spine	65%	N		N	N	Y
M06	Right Lung (upper lobe)	55%	Y	N	N	N	Y
M07	Liver (dome)	70%	N		N	N	Y
M08	Uterus	55%	N		N	N	Y
M09	Left Lung (upper lobe)	75%	Y	N	N	N	Y
M10	Carina (lymph node)	70%	N		N	N	Y
M11	Right Lung (lower lobe)	65%	Y	N	Y	Y	Y
M12	Liver (left lower lobe)	50%	Y	N	N	N	Y
M13	Liver (left lobe)	70%	N		N	N	Y
M15	Adrenal gland	40%	N		N	N	Y

Summary of the lesions collected during the autopsy of the index patient, cellularity assessment, imaging, clinical outcome, and sequencing techniques used. N, no; Y, yes.

Apico–basal forces exerted by apoptotic cells drive epithelium folding

Bruno Monier^{1,2*}, Melanie Gettings^{1,2*}, Guillaume Gay³, Thomas Mangeat^{1,2}, Sonia Schott^{1,2}, Ana Guarner⁴ & Magali Suzanne^{1,2}

Epithelium folding is a basic morphogenetic event that is essential in transforming simple two-dimensional epithelial sheets into three-dimensional structures in both vertebrates and invertebrates¹. Folding has been shown to rely on apical constriction^{2–7}. The resulting cell-shape changes depend either on adherens junction basal shift² or on a redistribution of myosin II^{3–5,7}, which could be driven by mechanical signals⁸. Yet the initial cellular mechanisms that trigger and coordinate cell remodelling remain largely unknown. Here we unravel the active role of apoptotic cells in initiating morphogenesis, thus revealing a novel mechanism of epithelium folding. We show that, in a live developing tissue, apoptotic cells exert a transient pulling force upon the apical surface of the epithelium through a highly dynamic apico-basal myosin II cable. The apoptotic cells then induce a non-autonomous increase in tissue tension together with cortical myosin II apical stabilization in the surrounding tissue, eventually resulting in epithelium folding. Together our results, supported by a theoretical biophysical three-dimensional model, identify an apoptotic myosin-II-dependent signal as the initial signal leading to cell reorganization and tissue folding. This work further reveals that, far from being passively eliminated as generally assumed (for example, during digit individualization⁹), apoptotic cells actively influence their surroundings and trigger tissue remodelling through regulation of tissue tension.

In different morphogenetic contexts, apoptosis has been shown to have an essential role in tissue folding; however, the cellular mechanisms involved remain mostly unknown^{10–12}. To characterize apoptosis-dependent folding, we focused on *Drosophila* leg epithelium morphogenesis, a process that has been shown to rely on local apoptosis¹¹ (Extended Data Fig. 1a, Supplementary Video 1). Interestingly, we discovered that leg disc folding follows a stereotypical sequence, with fold progression following the spreading of cell death, beginning in the most ventral part, then progressing laterally to end in the most dorsal region of the developing leg (Fig. 1a, b, Extended Data Fig. 1b). To unravel the link between apoptotic cells and fold formation, we first focused on apoptotic cell behaviour using live imaging. In the leg epithelium, apoptosis follows the classical morphological steps including cell shrinkage, membrane blebbing and fragmentation into apoptotic bodies (Extended Data Fig. 1c). Initially, apoptotic cells remain columnar and attached to their neighbours, as described previously¹³ (Extended Data Fig. 2a). Indeed, we noticed that adherens junction components (E-cadherin, α -catenin and β -catenin) accumulate below the apical surface of dying cells, forming an adhesion peak which coincides with local deformation of the apical surface of the surrounding epithelial cells (Fig. 1c, Supplementary Video 2). These observations prompted us to hypothesize either the presence of an apico-basal pulling force generated by the dying cells or, alternatively, a pushing force generated by the dying cell's neighbours. Therefore, we analysed myosin II dynamics. Interestingly, we detected an apico-basal acto-myosin cable-like structure (hereafter named 'cable') inside each dying cell (Extended Data Fig. 2b) that is formed just before the local deformation of the epithelium

surface (Fig. 1d, Supplementary Video 3). This myosin II cable is attached to the junctional structure described above (Fig. 1e). Remarkably, apical surface release coincides with myosin II cable and adhesion peak detachment from the apical surface as the dying cell fragments (Fig. 1c, d, Extended Data Fig. 2c, d). Furthermore, when apoptosis is inhibited, neither the myosin II cable nor the apical deformation are observed (Extended Data Fig. 2e), suggesting that the myosin II cable constitutes the cellular apoptotic machinery responsible for the transient deformation of the epithelium.

We then asked whether the apico-basal myosin II cable is a general characteristic of apoptotic epithelial cells. By analysing myosin II distribution in different *Drosophila* epithelial tissues, we revealed that an apico-basal myosin II cable also forms in apoptotic cells in other epithelia (Extended Data Fig. 2h). We further asked if this general property of apoptotic cells to generate an apico-basal myosin II cable is responsible for the local apical deformation observed around apoptotic cells. To test this, we induced ectopic apoptosis in the *Drosophila* wing which can be regarded as a naive tissue (as apoptosis normally occurs sporadically) and blocked myosin II function specifically in dying cells. While an apico-basal myosin II cable is formed in ectopic dying cells (Extended Data Fig. 2f), the local apical deformation around apoptotic cells is no longer visible when myosin II is inhibited (compare Fig. 1f to Fig. 1g, Extended Data Fig. 2g), indicating that the deformation strictly results from the myosin II dependent apoptotic force and is not generated by neighbouring cells. Together, these experiments demonstrate a fundamental intrinsic *in vivo* property of yet non-fragmented apoptotic cells, namely their ability to produce a myosin II dependent apico-basal pulling force capable of transiently deforming adjacent cells (Fig. 1h).

To determine how apoptotic cells control the reorganization of the remaining tissue, we characterized myosin II distribution and cell shape changes in the vicinity of apoptotic cells. Myosin II in this tissue is localized at the level of adherens junctions (Extended Data Fig. 3a, b). Interestingly, we observed that myosin II levels are increased along the apical membrane of apoptotic cell neighbours (Fig. 2a). Eventually, myosin II and F-actin apical stabilization is found throughout the whole fold domain where apoptosis takes place compared to the segment domain (Fig. 2b, Extended Data Fig. 3c, d) and is lost in absence of apoptosis (Fig. 2c and Extended Data Fig. 3e–g). Moreover, we observed that cells neighbouring apoptotic cells become progressively elongated and reduce their apical surface (Fig. 2d, Supplementary Video 4). Elongation then propagates from cell to cell, gradually spreading to the whole fold domain, thus generating a ring of stretched cells in which apoptosis occurs specifically (Extended Data Fig. 4a, b, Supplementary Video 5). These observations suggest that cell death is responsible for cell shape modification in the whole fold domain. We therefore compared cell shape dynamics of developing legs with or without apoptosis. In the control fold domain, cells elongate, decrease their apical surface and adopt a preferential orientation along the future fold (Fig. 2e). However, all these characteristics are lost when cell death is inhibited, demonstrating the essential role of apoptosis in determining cell morphology during

¹Université de Toulouse, UPS, LBCMCP, F-31062 Toulouse, France. ²CNRS, LBCMCP, F-31062 Toulouse, France. ³DamCB, Data Analysis and Modelling for Cell Biology, 13005 Marseille, France. ⁴Centro de Biología Molecular Severo Ochoa (C.S.I.C.-U.A.M.), Universidad Autónoma de Madrid, Nicolás Cabrera 1, Cantoblanco, 28049 Madrid, Spain.

*These authors contributed equally to this work.

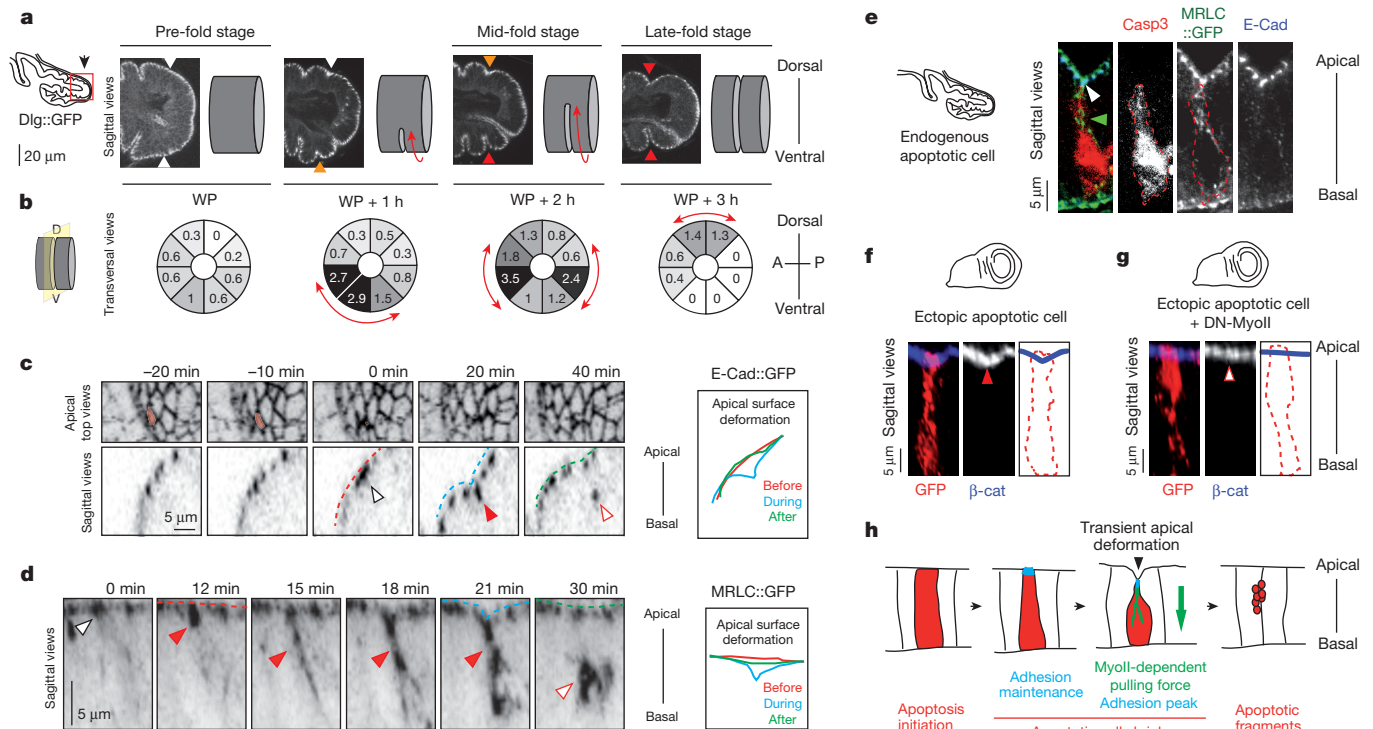


Figure 1 | Apoptotic cells exert a transient apico-basal force upon adjacent cells. **a**, Pupal leg disc from pre-fold stage (white pupae, WP) to late-fold stage (WP + 3 h). Arrowhead colours indicate fold progression ($n = 10, 7, 10$ and 10 , respectively). **b**, Average number of dying cells in the fold domain ($n = 10, 6, 10$ and 7 , respectively). A, anterior; P, posterior. **c**, **d**, E-cadherin–green fluorescent protein (E-Cad::GFP) (**c**, $n = 24$) and MRLC::GFP (**d**, $n = 21$) dynamics in apoptotic cells (in red in **c**) from pre-fold stage leg discs. Time 0 = collapse of apoptotic cell apical surface; the dotted line and arrowhead

colour codes indicate equivalent stages of the apoptotic process. **e**, Co-localization of the apoptotic myosin II cable (green arrowhead) with adherens junctions ($n = 19$, white arrowhead) stained with anti-E-Cad. **f**, **g**, Ectopic apoptotic cells with (**f**) or without (**g**) myosin II activity generated in the wing disc. DN, dominant negative form. Red and open arrowheads point at presence or absence of apical deformation, respectively (see quantification and genotypes in Extended Data Fig. 2g). **h**, Schematics of apoptotic cell dynamics.

folding (Fig. 2e and Extended Data Fig. 4c–e). Using laser ablation, we tested whether cell elongation and myosin II stabilization at the level of adherens junctions reveal an increase in tissue tension in the fold domain, as previously described in other tissues¹⁴. We found that the release of tension between vertices was indeed much higher in the fold domain where apoptosis takes place than in the segment domain (Extended Data Fig. 3h). In addition, when cell death was inhibited, tension

in the fold domain was significantly lower than in the control situation (Extended Data Fig. 3i). Altogether these results demonstrate that, during leg folding, apoptosis induces a non-autonomous effect throughout the fold domain leading to acto-myosin apical stabilization, a global enhancement of tissue tension and cell shape changes (Fig. 2f).

To test the role of apoptotic forces in folding, we constructed a physical model based on the two-dimensional vertex model¹⁵. In this model,

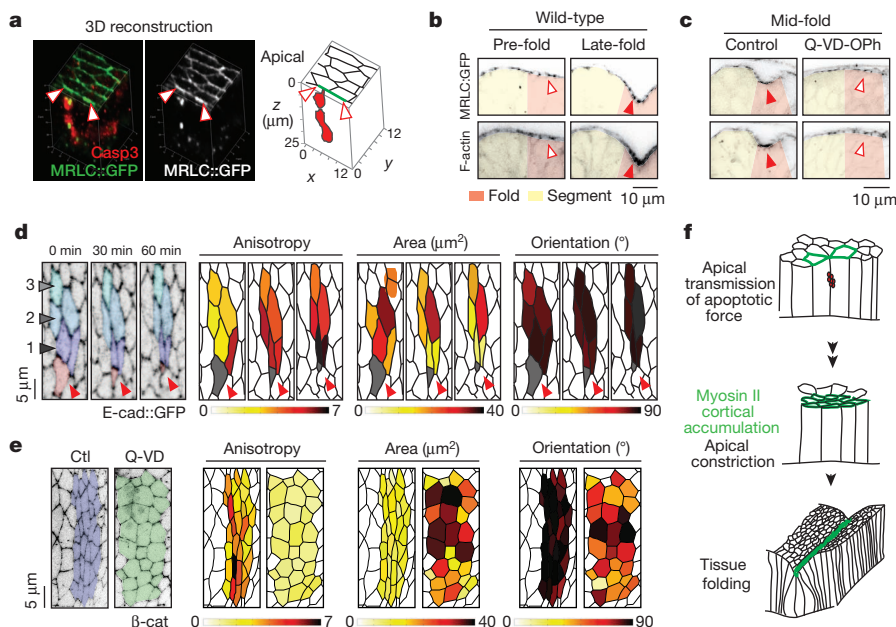


Figure 2 | The apoptotic force propagates to the neighbouring tissue. **a**, Leg disc three-dimensional reconstruction and schematics showing apical myosin II stabilization (arrowheads) in cells adjacent to a dead cell ($n = 11$). **b**, Leg discs from pre-fold stage ($n = 9$) and late-fold stage ($n = 7$). **c**, Leg discs incubated from pre-fold to mid-fold stages with DMSO (control, $n = 13$) or Q-VD-OPh (cell death inhibition, $n = 13$). Red and open arrowheads point to the presence or absence, respectively, of F-actin or myosin II stabilization in **b** and **c**. **d**, Leg disc time-lapse images and quantification of cell shape change parameters at the vicinity of an apoptotic cell (red arrowheads, $n = 18$). **e**, Quantification of cell shape change parameters of leg discs incubated with DMSO ($n = 7$) or Q-VD-OPh ($n = 8$) from pre-fold to mid-fold stages. **f**, Schematics of apoptotic force propagation.

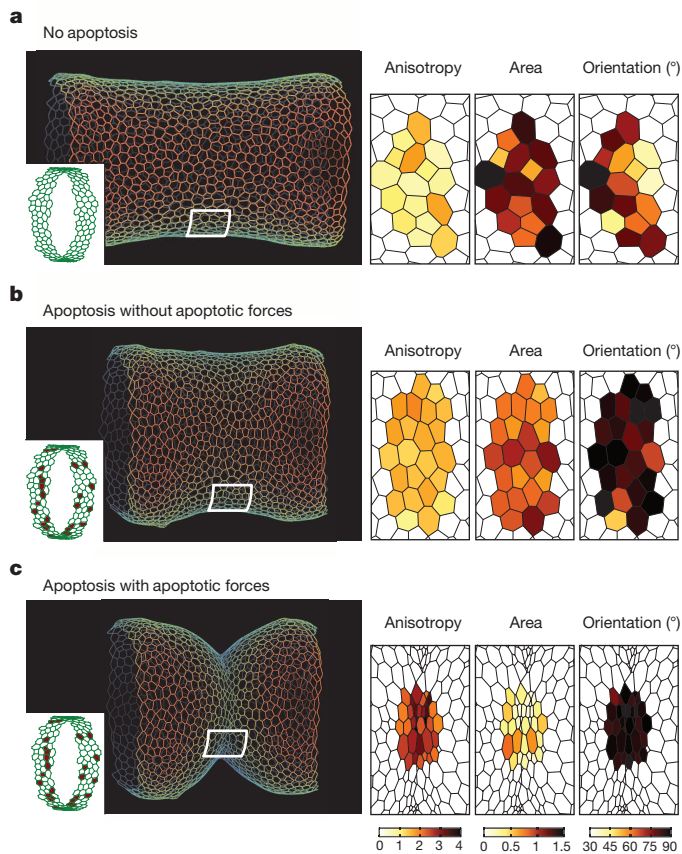


Figure 3 | Biophysical model of epithelium folding. **a–c**, *In silico* models showing apoptotic pattern (insets), whole tissue (left) and cell shape (right, including normalized area) in the absence of apoptosis (**a**) and after removing apoptotic cells from the fold domain without (**b**) and with (**c**) an apico-basal force generated by each dying cell and transmitted apically to neighbours (noted 'apoptotic forces').

three interactions are considered: cell elasticity dependent on the cell apical area, a contractility term dependent on the cell perimeter, and line tension dependent on apical junction length. To take into account all three dimensions of the leg epithelium, we have added an apico-basal tension to those interactions (Extended Data Fig. 5a), and have

considered cell elasticity as a function of cell volume rather than cell area while retaining the main characteristics of the original model (Extended Data Fig. 5b–e). Based on our observation of the developing leg disc, the leg tissue in our model is represented by a 50-cell circumferential cylinder, with the fold domain representing three rings of cells in which 30 cells are programmed to die following the pattern of apoptosis in the leg (compare the cell death pattern in Fig. 1b and Extended Data Fig. 6a). In the absence of the apoptotic-dependent forces, the model indicates that the simple disappearance of 30 cells from a continuous ring-like domain is not sufficient to induce cell shape reorganization and to create an invagination at the tissue level (compare Fig. 3a and b). However, if a transient apico-basal force is applied in each dying cell, an invagination response is observed all around the cylindrical tissue, albeit irregular (Extended Data Fig. 8a). Now, if an increase of apical contractility is applied in two rows of apoptotic neighbours (representing the non-autonomous increase in tissue tension), a cell shape reorganization is observed, although moderate (Extended Data Fig. 7a). Finally, if the transient pulling forces generated at a cellular scale are translated as increased contractility at the tissue scale we observe cell shape reorganization similar to that observed in the leg fold, along with regular and deeper folding (Fig. 3c). This shows that the added effect of both apoptotic forces is necessary and sufficient to induce folding *in silico*. Importantly, rising apoptotic cell number (Extended Data Fig. 6), apico-basal force strength (Extended Data Fig. 7) or the increase in apical contractility in apoptotic cells neighbours (Extended Data Fig. 8) leads to a gradation in cell and tissue shape changes and demonstrates the robustness of the model.

Furthermore, our model predicts that an apico-basal pulling force generated by sporadic apoptosis is not sufficient to modify tissue shape (Extended Data Fig. 6c, d). Moreover, the synergy of forces arising from several apoptotic cells concentrated in a restricted region appears necessary to generate a force strong enough to produce a fold (compare Extended Data Fig. 6f with Extended Data Fig. 6g). To test this *in vivo* we induced apoptosis ectopically in the wing pouch (that is, a flat tissue) and observed that a high concentration of apoptotic cells in a restricted region is indeed sufficient to drastically modify the shape of the epithelium through the creation of an ectopic fold (Fig. 4a, b, left and middle panels, and Extended Data Fig. 9a–b'). Consistently, no folding was observed in regions where only sporadic apoptotic cells were generated (Extended Data Fig. 9b'''). Importantly, this tissue bending coincides with an apical stabilization of myosin II (Fig. 4c, middle panel, Extended Data Fig. 9d–e') and F-actin (Extended Data Fig. 9c'). It also strictly relies on apoptotic myosin II since the expression of a dominant

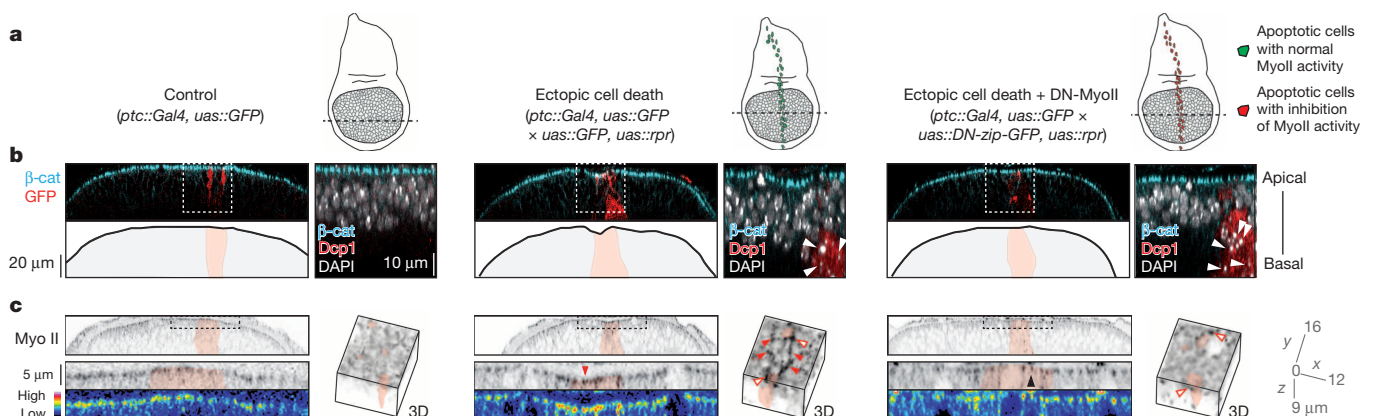


Figure 4 | Ectopic fold formation. **a**, Schematics of wing discs depicting the genetic contexts analysed in **b** and **c**. **b, c**, Wing discs in the absence (left panels) or presence (middle and right panels) of ectopic apoptosis in the *ptc* domain (pink), with (middle panels) or without (right panels) myosin II activity in dying cells, showing wing disc morphogenesis (large panels in **b**) and dying cell extrusion (close ups in **b**, white arrowheads) and myosin II distribution (stained by an anti-Sqh/MRLC, **c**; in sagittal sections, red and black arrowheads point to

the presence or absence, respectively, of myosin II apical stabilization). Close-up three-dimensional reconstructions are presented in **c** to visualize the different pools of apical myosin II and apoptotic cells; red and open arrowheads denote 'fold domain apical myosin II' and 'contractile ring of myosin II', respectively (see also Extended Data Fig. 9e–e'). **b**, $n = 8, 13$ and 10 ; **c**, $n = 6, 7$ and 8 in left, middle and right panels, respectively.

negative form of myosin II specifically in dying cells at the onset of apoptosis induction suppresses folding (Fig. 4a, b, right panels, and Extended Data Fig. 9a'', b''), whereas apoptotic cell extrusion from the epithelium remains normal (Fig. 4b, middle and right panels). Interestingly, the absence of fold in this context is concomitant with an absence of myosin II apical stabilization in the whole domain of ectopic apoptosis (although the myosin II contractile ring involved in dying cell extrusion¹⁶ is still present, see right panel in Fig. 4c and Extended Data Fig. 9e''), showing that this accumulation strictly depends on apoptotic myosin II (compare Fig. 4c middle panel with right panel, quantifications are in Extended Data Fig. 9d). Altogether, these data strongly suggest that the apoptotic force resulting from the synergy of numerous and patterned apoptotic events constitutes the primary signal leading to epithelium folding.

In this study, we elucidated a novel cellular mechanism of epithelial folding that relies on apoptosis. In *Drosophila* epithelia, cells are extremely columnar and apoptotic cells, in addition to the force generated by their extrusion from the epithelium as previously described in squamous epithelium¹⁷, create a myosin-II-dependent apico-basal intracellular pulling force. Based on previous publications showing that transient mechanical forces are sufficient to induce myosin II recruitment at the apical surface of an epithelium^{8,14}, we propose that apoptotic cells send a biomechanical signal to their neighbours, although we cannot exclude the myosin-II-dependent release of a molecular signal. The apoptotic signal triggers non-autonomous myosin II recruitment at the level of adherens junctions of neighbouring cells. The synergy of several apoptotic cells leads to a redistribution of myosin II, increased tension, apical constriction in the entire fold domain and subsequent tissue folding (Extended Data Fig. 10). Taken together, these results reveal that through the regulation of tissue tension, apoptotic cells can actively control tissue remodelling.

Interestingly, apoptosis-dependent folding has also been described in vertebrates during neural tube bending. This bending is an important step in neural tube closure and its failure can lead to spina bifida phenotypes^{12,18}. Therefore, it would be interesting to test if the cellular mechanisms described here are conserved in vertebrates, thus generalizing the mechanism of apoptosis-dependent epithelium folding.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 10 December 2013; accepted 9 December 2014.

Published online 21 January 2015.

- Davidson, L. A. Epithelial machines that shape the embryo. *Trends Cell Biol.* **22**, 82–87 (2012).
- Wang, Y. C., Khan, Z., Kaschube, M. & Wieschaus, E. F. Differential positioning of adherens junctions is associated with initiation of epithelial folding. *Nature* **484**, 390–393 (2012).
- Martin, A. C., Kaschube, M. & Wieschaus, E. F. Pulsed contractions of an actin-myosin network drive apical constriction. *Nature* **457**, 495–499 (2009).

- Roh-Johnson, M. *et al.* Triggering a cell shape change by exploiting preexisting actomyosin contractions. *Science* **335**, 1232–1235 (2012).
- Sherrard, K., Robin, F., Lemaire, P. & Munro, E. Sequential activation of apical and basolateral contractility drives ascidian endoderm invagination. *Curr. Biol.* **20**, 1499–1510 (2010).
- He, B., Dubrovinski, K., Polyakov, O. & Wieschaus, E. Apical constriction drives tissue-scale hydrodynamic flow to mediate cell elongation. *Nature* **508**, 392–396 (2014).
- Brodu, V. & Casanova, J. The RhoGAP *crossveinless-c* links *trachealess* and EGFR signaling to cell shape remodeling in *Drosophila* tracheal invagination. *Genes Dev.* **20**, 1817–1828 (2006).
- Pouille, P. A., Ahmadi, P., Brunet, A. C. & Farge, E. Mechanical signals trigger myosin II redistribution and mesoderm invagination in *Drosophila* embryos. *Sci. Signal.* **2**, ra16 (2009).
- Montero, J. A. & Hurlé, J. M. Sculpting digit shape by cell death. *Apoptosis* **15**, 365–375 (2010).
- Lohmann, I., McGinnis, N., Bodmer, M. & McGinnis, W. The *Drosophila* *Hox* gene *deformed* sculpts head morphology via direct regulation of the apoptosis activator *reaper*. *Cell* **110**, 457–466 (2002).
- Manjón, C., Sánchez-Herrero, E. & Suzanne, M. Sharp boundaries of Dpp signalling trigger local cell death required for *Drosophila* leg morphogenesis. *Nature Cell Biol.* **9**, 57–63 (2007).
- Yamaguchi, Y. *et al.* Live imaging of apoptosis in a novel transgenic mouse highlights its role in neural tube closure. *J. Cell Biol.* **195**, 1047–1060 (2011).
- Lubkov, V. & Bar-Sagi, D. E-cadherin-mediated cell coupling is required for apoptotic cell extrusion. *Curr. Biol.* **24**, 868–874 (2014).
- Fernandez-Gonzalez, R., Simoes, S. M., Röper, J. C., Eaton, S. & Zallen, J. A. Myosin II dynamics are regulated by tension in intercalating cells. *Dev. Cell* **17**, 736–743 (2009).
- Farhadifar, R., Röper, J. C., Aigouy, B., Eaton, S. & Jülicher, F. The influence of cell mechanics, cell-cell interactions, and proliferation on epithelial packing. *Curr. Biol.* **17**, 2095–2104 (2007).
- Rosenblatt, J., Raff, M. C. & Cramer, L. P. An epithelial cell destined for apoptosis signals its neighbors to extrude it by an actin- and myosin-dependent mechanism. *Curr. Biol.* **11**, 1847–1857 (2001).
- Toyama, Y., Peralta, X. G., Wells, A. R., Kiehart, D. P. & Edwards, G. S. Apoptotic force and tissue dynamics during *Drosophila* embryogenesis. *Science* **321**, 1683–1686 (2008).
- Yamaguchi, Y. & Miura, M. How to form and close the brain: insight into the mechanism of cranial neural tube closure in mammals. *Cell. Mol. Life Sci.* **70**, 3171–3186 (2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank C. Benassayag, E. Farge, Y. Gachet, T. Lecuit, P.-F. Lenne, F. Payre, E. Sanchez-Herrero, B. Sanson and S. Tournier for comments on the manuscript, T. de Paula Peixoto for providing the graph-tool library, D. Kiehart, H. Steller, X. Wang, R. Ward, BDSC and DSHB for stocks and reagents, the TRI platform for imaging facilities and M. Aguirrebengoa for helping us with statistics. The Suzanne laboratory is supported by grants from the Agence Nationale de la Recherche (ANR), Fondation de la Recherche et de l'Innovation Thérapeutique en Cancérologie (RITC) and the University of Toulouse.

Author Contributions M.G., B.M. and M.S. have done the experiments in flies with the help of S.S. and A.G.; G.G. designed the simulation model. T.M. set up cell shape extraction, statistics and analysis in Matlab and performed the laser ablation experiments with the help of B.M.; M.S. supervised the project.

Author Information The source code for the model is released under the GNU General Public Licence and is available on GitHub (<https://github.com/glyg/leg-joint> and <http://dx.doi.org/10.5281/zenodo.13386>). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.S. (magali.suzanne@univ-tlse3.fr) for biology and materials and to G.G. (guillaume@damcb.com) for modelling.

METHODS

Fly stocks and genetics. The fluorescent reporters used are the following: E-Cad-KI(GFP)¹⁹, *ubi::E-Cad::GFP*²⁰, *uas::alpha-catenin-TagRFP*²¹, *uas::SCAT3* (FRET reporter of caspases activity²²), *sqh[AX3];sqh::sqhGFP[40]* and *sqh[AX3];sqh::sqhGFP[42]* (MRLC::GFP)²³, *w;ap::Gal4,arm::arm-GFP;uas::mCD8-Cherry* (generated using Bloomington stocks), *w,Dlg1-GFP* (CC01936, from Flytrap), *Dll::Gal4 MD23, zfh2 LP30::Gal4* and *UAS::Diap1* are described in Flybase, *UAS::p35* (insertions in chromosome II and III) come from Bloomington.

Stocks for ectopic cell death induction are *y,w,HS::flp;act>y+>Gal4,uas::GFP* and *w;ptc::Gal4, uas::GFP; tub::Gal80ts* (gifts from C. Benassayag), *y,w,HS::flp; act5C>CD2> Gal4* (generated using Bloomington stocks), *w;uas::lifesactGFP;uas::reaper* (gifts from X. Wang and from Flybase), *w; uas::hid(4),uas::DN-zip::GFP* (gifts from H. Steller and D. Kiehart), *w; uas::DN-zip::GFP; uas::rpr*. Briefly, the progeny of crosses of interest were grown on standard medium at 25 °C. Third instar larvae were heat shocked for 15–20 min at 38 °C and transferred to 29 °C for 5.5 h–6 h before dissection. Following this treatment, a reproducible bias was observed, with clones essentially following the dorsal-ventral boundary in half of the wing pouch. When using *ptc::Gal4* to induce cell death, crosses were performed at 18 °C, the progeny was transferred at least 5.5 h to 30 °C before dissection.

Immunostainings. Primary antibodies obtained from Developmental Studies Hybridoma Bank were: rat anti-E-Cad (DCAD2, 1:50), mouse anti-Arm (N2 7A1, 1:5) and mouse anti-Dlg (4F3, 1:200). Rabbit anti-cleaved Caspase 3 (9661, 1:100) and anti-cleaved Dcp1 (9578, 1:200) were obtained from Cell Signaling Technologies and chicken anti-beta-Gal (GTX77365, 1:1,000) was obtained from GeneTex. Mouse anti-Sqh (MRLC) and Guinea pig anti-Sqh1P (1:1,000) were gifts from R. Ward. Staining of the actin cytoskeleton was achieved using phalloidin-Rhodamine (1:200, Invitrogen) or phalloidin-Alexa647 (1:100, Interchim). Secondary antibodies coupled to Alexa-488 or-555 were obtained from Fisher Scientific and diluted 1:200 while secondary antibodies coupled to Cy5 were obtained from Jackson Laboratories and diluted 1:50. Briefly, for immunostainings, imaginal tissues were fixed using paraformaldehyde (PFA) 4% diluted in PBS 1×. Samples were washed and saturated using PBS 1×, Triton X-100, 0.3% BSA 1% (BBT). Primary antibodies were diluted in BBT and incubated overnight at 4 °C. Next, samples were washed and saturated in BBT, incubated with secondary antibodies (and phalloidin if required to stain F-actin) as indicated above, and subsequently washed with PBS 1×, Triton X-100, 0.3%. Samples were mounted in Vectashield containing DAPI (Vectors laboratories) and analysed under a Zeiss LSM710 laser scanning microscope. A similar protocol was followed for immunostainings on embryos except that fixation was performed for 5 min in heptane:formaldehyde 37% (1:1). For E-cadherin and MyoII, embryos were devitelinated manually and stained immediately. Note that in order to preserve wing morphology in ectopic cell death experiments; dissections were performed in Schneider medium, followed by fixation with 4% PFA diluted in Schneider medium.

Time-lapse imaging. Leg discs were dissected in Schneider's insect medium (from Sigma Aldrich) supplemented with 2% FCS and 0.5% penicillin-streptomycin. Ecdysone (from Sigma Aldrich, 20-hydroxyecdysone H5142) was stored in a stock solution of 200 µg ml⁻¹ at -20 °C and added to meet a final concentration of 2 µg ml⁻¹. For *in vivo* imaging, leg discs were transferred on a coverslip in 15 µl of the above medium complemented with methyl cellulose (from Sigma-Aldrich) at a final concentration of 2.5% to obtain a more viscous medium²⁴. Spacers (Secure-Seal Imaging Spacers 0.12 depth from Sigma-Aldrich) were added between the coverslip and an air-permeable membrane (Lumox 25 from Sarstedt), to avoid compression of the tissue, and halocarbon oil was added on the sides of the spacer to protect from dehydration. Note that forceps and scissors, as well as air-permeable membrane, were washed with ethanol before dissection. The membrane is rinsed with water and dried before use.

Before imaging, dissected leg discs may be incubated for 30 min with acridine orange (final concentration 0.1 µg ml⁻¹) to reveal dying cells or with the red fluorescent lipid-binding dye FM4-64 (final concentration 36 µM). FM4-64 was also added at a concentration of 18 µM in the imaging medium.

Imaging was essentially performed under an inverted laser scanning LSM710 confocal (Zeiss). Under our conditions, *ex vivo* leg development reproduces morphologic stages of pupal leg development characterized on fixed tissues (including leg evagination, local folding and the pattern of cell death), albeit at a lower speed (around 2 times).

Importantly, we found that image stacks of 30–40 µm have to be taken every 3 min with optimal sectioning (0.438 µm for a 40× objective with 1.3 aperture and the pinhole set to 1 AU) in order to spot apico-basal myosin II cables and adhesion peaks as they are very dynamic structures.

Note that apoptotic cells were identified using the caspase FRET sensor (SCAT3²², see section 'Post-imaging analysis' for details), acridine orange²⁵ or FM4-64 that strongly labels apoptotic bodies.

Ex vivo culture and drug treatment. Q-VD-OPH is a broad spectrum caspase inhibitor which binds to the active site of activated proteases, more efficiently than the commonly used Z-VAD caspase inhibitor and is described as non-toxic even at extremely high concentrations²⁶. Discs from white pupae (WP) were dissected in complemented Schneider medium as described above and incubated from pre-fold to mid-fold stage with either Q-VD-OPH (from R&D systems; final concentration 500 µM) or DMSO (from Sigma Aldrich, final concentration 0.5% in Fig. 2e and 0.06% in other experiments) in Fig. 2c (incubation of 6h) or Fig. 2e and Extended Data Fig. 4d (incubation of 4 h). Q-VD-OPH was stored in a stock solution of 10 mM in either DMSO 100% (Fig. 2e) or in Schneider + 1.25% DMSO (other experiments). Throughout this study, we focused on the t4-t5 fold, the only fold that is exclusively formed during the pupal stage and for which progression can be easily followed due to disc evagination. At WP stage, formation of the t4-t5 fold is not initiated yet, while other tarsal folds are partially formed. Hence, in this particular context of apoptosis inhibition, we observe, as expected, a slight perturbation of the formation of t3-t4 and strongly perturb the t4-t5 fold.

Post imaging analysis. SCAT3 FRET probe was previously described²². It is an indicator of caspase-3 activation using fluorescence resonance energy transfer (FRET) between an enhanced cyan fluorescent protein (the donor) and an enhanced yellow fluorescent protein (the acceptor) separated by a caspase cleavage site. FRET images of live leg discs were acquired with a Zeiss LSM710 microscope. A 458 nm laser was used to excite the sample. Cyan fluorescent protein (CFP) and yellow fluorescent protein (YFP) emission signals were collected through channel I (470–510 nm) and channel II (525–600 nm), respectively. CFP and YFP images were acquired simultaneously. Sequential acquisition of CFP and YFP channels in alternative orders were tested and gave the same result as simultaneous acquisition. CFP and YFP images were processed by ImageJ software. A background region of interest was subtracted from the original image. Gaussian smooth filter was then applied to both channels. The final ratio image (YFP/CFP signal) was generated in ImageJ program.

Zen software (Zeiss) was used to generate three-dimensional reconstruction and sagittal views of tissues. Images were processed in Adobe Photoshop CS5 or ImageJ.

Automated image analysis and quantifications of cell shape changes. Matlab, DipImage, Cell profiler and ICY were used to make automatic segmentation for quantification of cell shape in Fig. 2e and Extended Data Fig. 4d). Three-dimensional median filter was performed from confocal images before segmentation. A Z projection based on the maximum intensity from each Z stack was applied. An adaptive threshold was made to the Z projection images to define the outline of cells based on the Otsu method. The local maximum of the "distance" function was used to find the centre of each cell to produce a resulting binary image. The distance transform is then used on the binary image to calculate the distance from cell membrane. The local maxima of the distance transform was calculated and were selected as seeds. For cells with high anisotropy the calculation of local maximum induced some errors and was manually corrected. The coordinate of each local maximum was used like seed to use a watershed algorithm function in Matlab and obtain a label matrix allowing the quantification of each cell object. The cells situated out of the domain of interest were rejected. To perform statistics we calculated the maximum and the minimum vertical coordinates based on the coordinates of the centre of outlined cells. Then 12 domains were defined based on this coordinate. The calculation of anisotropy was defined as the ratio between the lengths of the principal axes of an ellipsis fitting the cell contour.

Measurements of actomyosin levels. To quantify junctional acto-myosin (Extended Data Fig. 3c). For each leg disc, acto-myosin intensity was measured using ImageJ by drawing a circular region of interest of 1 µm² in 8 individual junctions in the fold domain (t4-t5) and in 8 individual junctions in the segment domain (t4). The measurements done in the fold were then normalized with the mean value of acto-myosin intensity in the segment domain independently in each leg disc (to insure the presence of an internal control). As a consequence, the values represent the differential accumulation of junctional acto-myosin in the fold compared to the segment domain.

To quantify acto-myosin per surface unit (Extended Data Fig. 3d, e, g). For each leg disc, acto-myosin intensity was measured using ImageJ by drawing a segmented line either in the fold (t4-t5) or in the segment (t4) domain. These values were divided by the length of the line and each measure in the fold was then normalized by the corresponding segment domain value (to insure the presence of an internal control). As a consequence, the values represent the differential accumulation per surface unit of acto-myosin in the fold compared to the segment.

Statistics. To calculate P-values, we used the non-parametric Wilcoxon rank sum test (also called Mann and Whitney test) since samples do not follow a normal distribution and do not have equal variances in Extended Data Fig. 3c–e, g, Extended Data Fig. 4d and Extended Data Fig. 9d. The null hypothesis is that the measures are samples from continuous distributions with equal medians. The test considers that the samples are independent and in Extended Data Fig. 4d, it takes into account

that the control and Q-VD-OPh can have different lengths. For cell shape characteristics (anisotropy, area and orientation), values are represented by box plot (the red line represent the median). For acto-myosin level quantifications, values are represented as mean values with standard error bars.

Photo-ablation experiments. Laser-ablation experiments in prepupal leg imaginal discs were performed with a pulsed Q-switched microchip double-frequency Nd:YAG laser ($\lambda = 532$ nm, 550 ps, 7 kHz, 3.5 μ J per pulse). The laser beam was focused through a high numerical aperture oil-immersion lens ($\times 60$ C-Apochromat NA 1.4, Leica) to produce an experimental beam waist around 600 nm in the focal plane. Photo-disruption was produced in the focal plane due to the plasma generation in the middle of the adherens junctions after a total of 21 laser pulses of 3 kHz. A pair of Galvanometric mirrors were conjugated to the black focal plane of the microscope objective lens to stir the ablation beam and allow simultaneous ablation of two junctions. We used an inverted microscope (Leica DMI6000B) with wide field illumination to produce fluorescent live imaging of adherens junctions labelled with arm::GFP fusion protein. Adherens junctions were placed in the centre of the field to produce a better reproducibility. METAMORPH coupled to ILAS software controlled the laser and the microscope (RoperScientific SA). The Images were taken every second over a period of 30 s. Note that for laser ablations, in absence of cell death (incubation with Q-VD-OPh, Extended Data Fig. 3i), expression of a membrane-bound cherry protein under the control of the apterous driver (w;ap::Gal4, arm::arm-GFP;uas::mCD8-Cherry; for expression pattern, see Extended Data Fig. 4c) was used to visualize the fold domain since cell shape changes do not occur in these conditions.

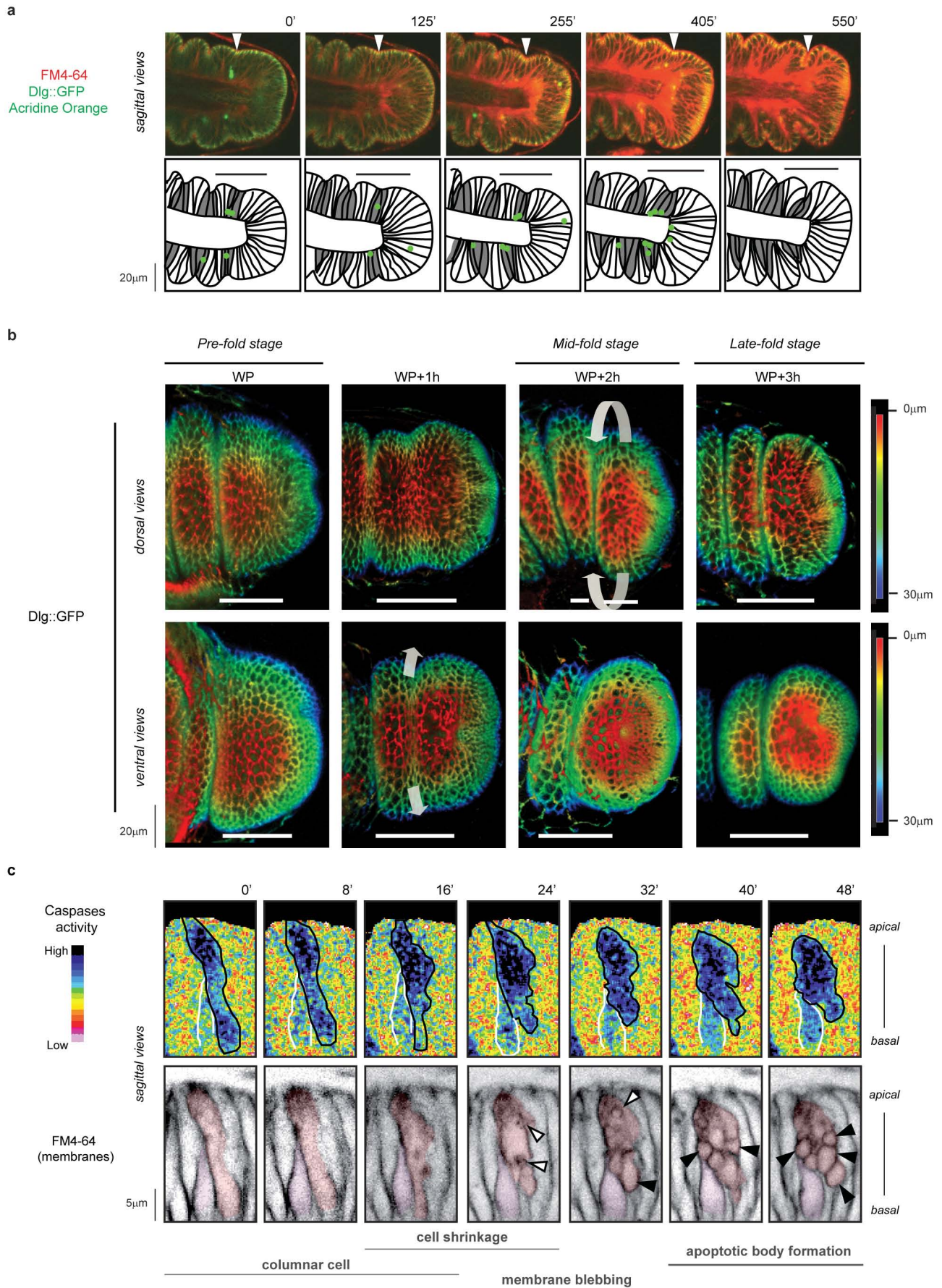
Modelling. Our apical junction network model is a generalization to three-dimensional of the Farhadifar *et al.* model¹⁵. In order to allow non-planar interactions, the surface elasticity is replaced by a volume elasticity. The apico-basal interaction is modelled as an energy proportional to cell height, with the same linear form as the apical line tension. The initial epithelium is represented as an hexagonal lattice of cells over a cylinder. Next, each cell follows a division process with a random division plane orientation. The order of cell divisions on the epithelium is chosen randomly, uniformly across the epithelium. Division is modelled by first increasing the cell equilibrium volume, finding the local energy minimum and second by dividing the cell and finding the new local energy minimum. Then, apoptotic cells are chosen randomly in a region around the centre of the epithelium, with a biased distribution reproducing the *in vivo* cell death pattern. Apoptosis is performed by gradually diminishing the cell preferred volume, increasing its contractility and rising the apico-basal tension term (ten steps are performed). At each step, the local energy is minimised by a gradient descent strategy. After the first step of the first apoptotic cell is performed, the second cell starts its apoptosis, and so on. Once ten steps are performed for one cell, it is removed from the tissue, such that no centripetal force remains at that point. When removing the apoptotic cell, its neighbours are rearranged through a series of type I intercalations (see Farhadifar *et al.*¹⁵). To model acto-myosin activity increase in the neighbouring cells, contractility is increased in the neighbouring cells by a factor the amplitude of which decreases exponentially with the distance to the apoptotic cells. After a cell has been eliminated and its neighbours rearranged, the whole tissue is brought back to equilibrium by successively computing the local energy minimum for all the cells of the tissue in a random order. Technically, the epithelium is described as an oriented

graph in the graph-tool library²⁸ (<http://graph-tool.skewed.de>) in the Python programming language. Energy minimisation is performed with the scipy library²⁹ (<http://scipy.org>), using the Broyden, Fletcher, Goldfarb, and Shanno bound constrained minimisation algorithm provided by this library. The source code for the model is released under the GNU General Public Licence and is available on github at <https://github.com/glyg/leg-joint> and <http://dx.doi.org/10.5281/zenodo.13386>. Extensive details on the numerical method as well as a complete derivation of the gradient can be found at this address.

The apical contractility and linear tension parameters Γ and Λ were chosen so that the relation between cell apical area and number of neighbours follows the same linear increase as observed by Farhadifar *et al.*¹⁵ in the wing disc (compare Fig. 2g of Farhadifar *et al.*¹⁵ with Extended Data Fig. 5b of the present work). We explored a range of apoptotic cell number, apico-basal force amplitude and contractility increase values (Extended Data Figs 6, 7, 8) and observed that cell morphology in the fold is best reproduced with amplitudes of 1 Γ and 1 Λ . The fact that they are in the same order of magnitude as the apical interactions gives us confidence that the chosen parameters are realistic. Furthermore, our tissue has a consistent morphology with or without apoptosis, increased apical contractility or apico-basal force, meaning that the balance of forces is coherent and equilibrated.

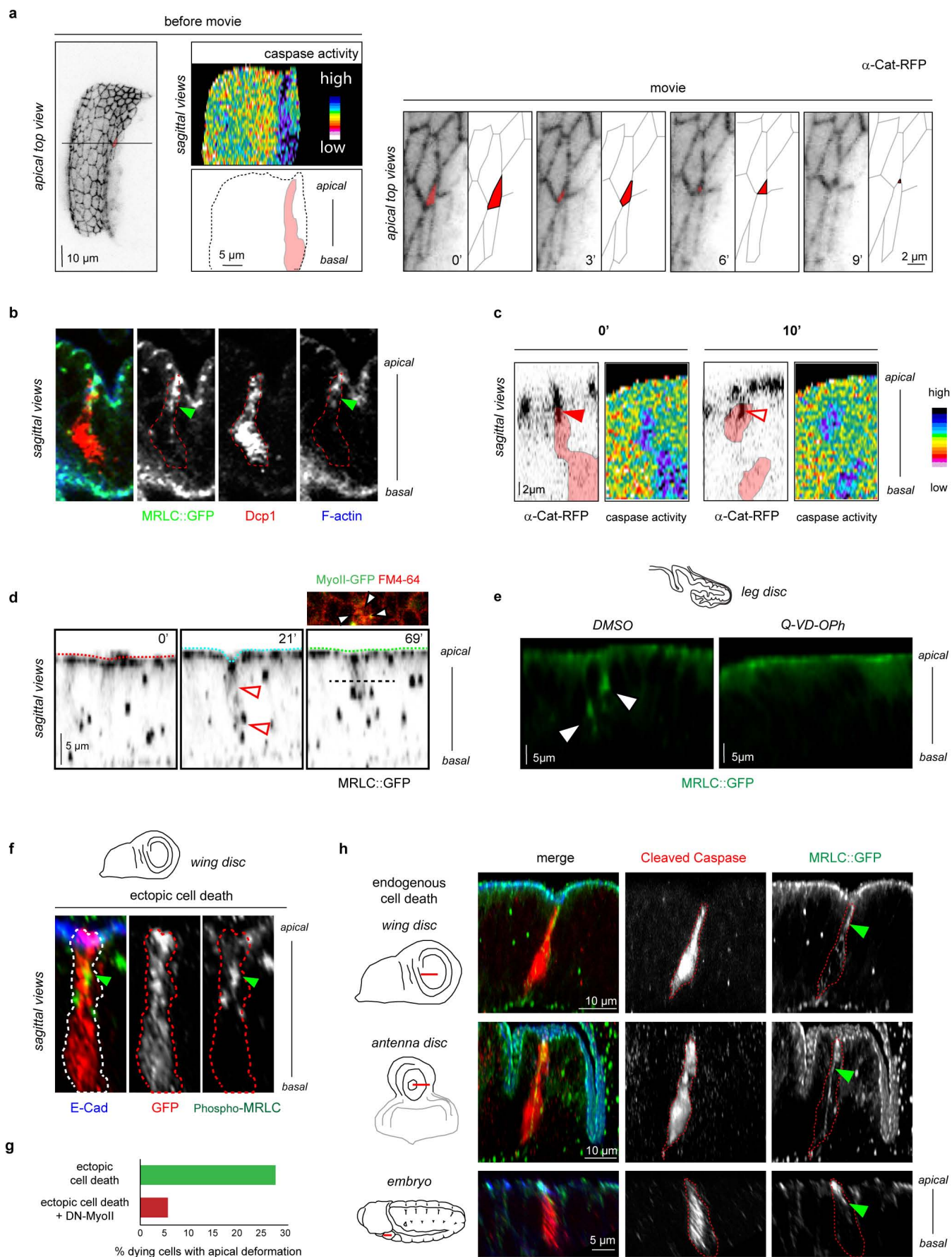
For cell shape analysis, a sector centred 30° above the ventral most part of the cylindrical epithelium was extracted. Radius and area were normalized with respect to their average values on the simulated epithelium in the absence of apoptosis.

19. Huang, J., Zhou, W., Dong, W., Watson, A. M. & Hong, Y. From the Cover: Directed, efficient, and versatile modifications of the *Drosophila* genome by genomic engineering. *Proc. Natl Acad. Sci. USA* **106**, 8284–8289 (2009).
20. Oda, H. & Tsukita, S. Real-time imaging of cell-cell adherens junctions reveals that *Drosophila* mesoderm invagination begins with two phases of apical constriction of cells. *J. Cell Sci.* **114**, 493–501 (2001).
21. Ishihara, S. & Sugimura, K. Bayesian inference of force dynamics during morphogenesis. *J. Theor. Biol.* **313**, 201–211 (2012).
22. Takemoto, K., Nagai, T., Miyawaki, A. & Miura, M. Spatio-temporal activation of caspase revealed by indicator that is insensitive to environmental effects. *J. Cell Biol.* **160**, 235–243 (2003).
23. Royou, A., Sullivan, W. & Karess, R. Cortical recruitment of nonmuscle myosin II in early syncytial *Drosophila* embryos: its role in nuclear axial expansion and its regulation by Cdc2 activity. *J. Cell Biol.* **158**, 127–137 (2002).
24. Aldaz, S., Escudero, L. M. & Freeman, M. Live imaging of *Drosophila* imaginal disc development. *Proc. Natl Acad. Sci. USA* **107**, 14217–14222 (2010).
25. Abrams, J. M., White, K., Fessler, L. I. & Steller, H. Programmed cell death during *Drosophila* embryogenesis. *Development* **117**, 29–43 (1993).
26. Caserta, T. M., Smith, A. N., Gultice, A. D., Reedy, M. A. & Brown, T. L. Q-VD-OPh, a broad spectrum caspase inhibitor with potent antiapoptotic properties. *Apoptosis* **8**, 345–352 (2003).
27. Guarner, A. *et al.* The zinc finger homeodomain-2 gene of *Drosophila* controls Notch targets and regulates apoptosis in the tarsal segments. *Dev. Biol.* **385**, 350–365 (2014).
28. Peixoto, P. T. The graph-tool python library. *figshare* <http://dx.doi.org/10.6084/m9.figshare.1164194> (2014).
29. van der Walt, S., Colbert, S. C. & Varoquaux, G. The NumPy Array: a structure for efficient numerical computation. *Comput. Sci. Eng.* **13**, 22–30 (2011).



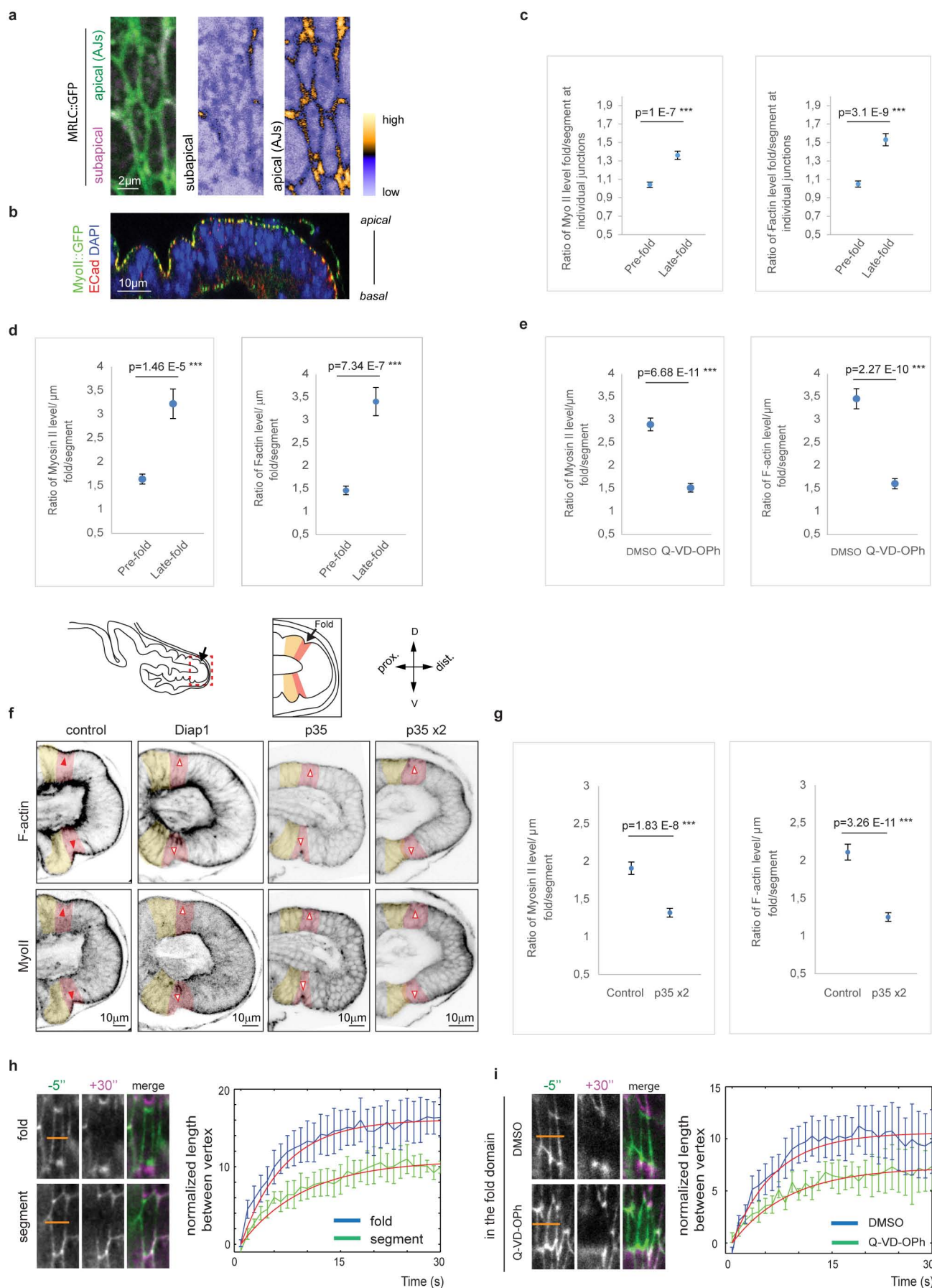
Extended Data Figure 1 | Spatio-temporal pattern of fold formation and hallmarks of apoptosis. This figure is associated with Fig. 1. **a**, Time-lapse images and schematics of the distal region of a Dlg::GFP leg disc from pre-fold stage (WP) showing the progression of the t4-t5 fold (white arrowheads). Acridine orange was used to stain dying cells (green) and FM4-64 to stain membranes (red). Note the presence of apoptotic cells in the fold region (coloured in grey in the schematics below, $n = 27$). The t4-t5 domain is indicated by a black line on the schematics. **b**, Three-dimensional reconstruction of the distal region of Dlg::GFP pupal legs undergoing fold morphogenesis. The colour code indicates tissue depth. Images show legs at different stages of development, from WP to WP + 3 h. Throughout this study, we have focused on the t4-t5 fold, the only one that is exclusively formed during pupal stage and for which progression can be easily followed due to disc evagination. For each time point, top and bottom panels show dorsal and

ventral views of leg discs respectively, and the t4-t5 domain is indicated by a white line. Note that the fold is initiated in the most ventral part of the leg, then progresses laterally (arrows) to end in the most dorsal part of the leg ($n = 10$ for WP, $n = 7$ for WP + 1 h, $n = 10$ for WP + 2 h, $n = 10$ for WP + 3 h). **c**, High magnification images from a time-lapse video during apoptosis showing caspase activity, revealed by the FRET construct SCAT3 (top), and the outline of cell membranes, revealed by FM4-64 staining (bottom) ($n = 19$). These images illustrate that the classical apoptotic stages, including shrinkage, blebbing (hollow arrowheads) and fragmentation (black arrowheads), are recapitulated in the developing *Drosophila* leg epithelium. Black outline (top) and red false-colour (bottom) highlight the apoptotic cell. Another apoptotic cell (outlined in white and coloured in pink) has also just turned on the apoptotic pathway. Note that in both cases the apoptotic pathway is turned on before visible morphological change.



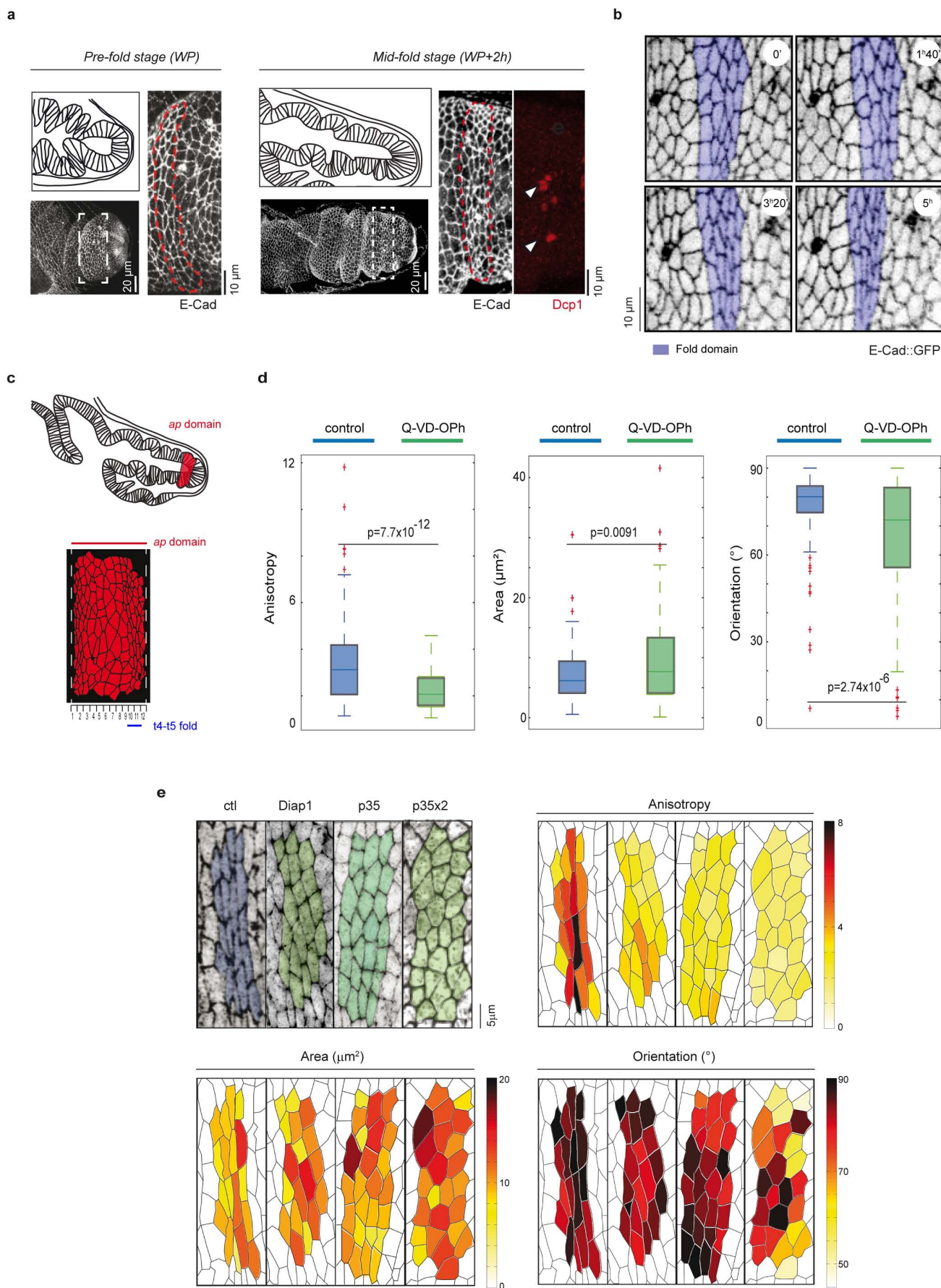
Extended Data Figure 2 | Adherens junctions and acto-myosin cytoskeleton dynamics in apoptotic cells. This figure is associated with Fig. 1. **a**, Time lapse images of a pupal leg disc expressing α Cat-TagRFP in the *ap* domain (which encompasses the t4-t5 fold) and the FRET construct SCAT3 to reveal caspase activity and thus visualize apoptotic cells ($n = 10$). The apoptotic cell outline is visible on the sagittal section and represented on the scheme. The position of the sagittal section is indicated by a black line on the apical top view to the left. Note the reduction of the apical surface of the apoptotic cell (apical top views on the right). The apoptotic cell is highlighted in red on both panels. **b**, Acto-myosin cable (green arrowhead) observed in a single apoptotic cell from a pupal leg disc ($n = 6$). MRLC::GFP is in green, F-actin (stained with phalloidine) is blue and cleaved Dcp1 is red. **c**, Sagittal sections of an apoptotic cell (visualized with the FRET construct SCAT3) before and after fragmentation, illustrating the apoptotic adhesion peak ($n = 27$, red arrowhead) and its splitting up (hollow red arrowhead) when the apoptotic cell detaches at fragmentation. **d**, Time-lapse images of a MRLC::GFP pupal leg disc showing apical surface release upon apoptotic cell fragmentation (time-point 69 min, $n = 12$). Fragmentation is clearly visible in the *z* section (top) showing membrane staining of the apoptotic fragments with FM4-64 (white arrowheads) (the position of the *z* section is indicated by a black dotted line in the sagittal view). The myosin II cable is indicated by hollow open red arrowheads in sagittal sections. **e**, MRLC::GFP leg discs incubated in either

DMSO (control) or Q-VD-OPH (cell death inhibition), showing the absence of the apico-basal myosin II cable in the absence of cell death (right, $n = 0$ out of 30). Arrowhead points out to the myosin II cable in the control ($n = 14$ out of 30). **f**, A single apoptotic cell (same cell shown in Fig. 1f, identified by GFP expression, red) generated by ectopic expression of *reaper* in the wing disc (from *y,w,hs::flp; act-*frt*-y+ -*frt*-Gal4, uas::GFP / uas::lifeactGFP; uas::rpr* larvae). The green arrowhead points out to the myosin II cable stained by anti-phospho-Sqh/MRLC (green, $n = 11$). The dotted red line outlines the dying cell. **g**, Percentage of individual apoptotic cells (with or without myosin II activity) with apical deformation. This quantification is associated with Fig. 1f, *g* ($n = 68$ and $n = 106$, respectively). Genotype for ectopic cell death is *y,w,hs::flp; act-*frt*-stop-*frt*::Gal4, uas::GFP / uas::lifeactGFP; uas::rpr* and genotype for ectopic cell death without myosin II activity specifically in apoptotic cells is *y,w,hs::flp; uas::DN-*zip*::GFP, uas::hid / act-*frt*-stop-*frt*::Gal4*. Note that, in the latter condition, myosin II activity is maintained in the neighbouring cells. **h**, Sagittal sections close-ups of MRLC::GFP wing and antennal discs from 3rd instar larvae and a MRLC::GFP stage 11 embryo showing that an apico-basal structure of myosin II (green arrowheads) is formed in dying cells in each of these different tissues ($n = 6, 7$ and 7 , respectively). Cleaved Dcp1 (activated caspase) is in red, E-Cad in blue (wing) and Arm/ β Cat in blue. The region shown in each close up is indicated by a red line on schematics on the left.



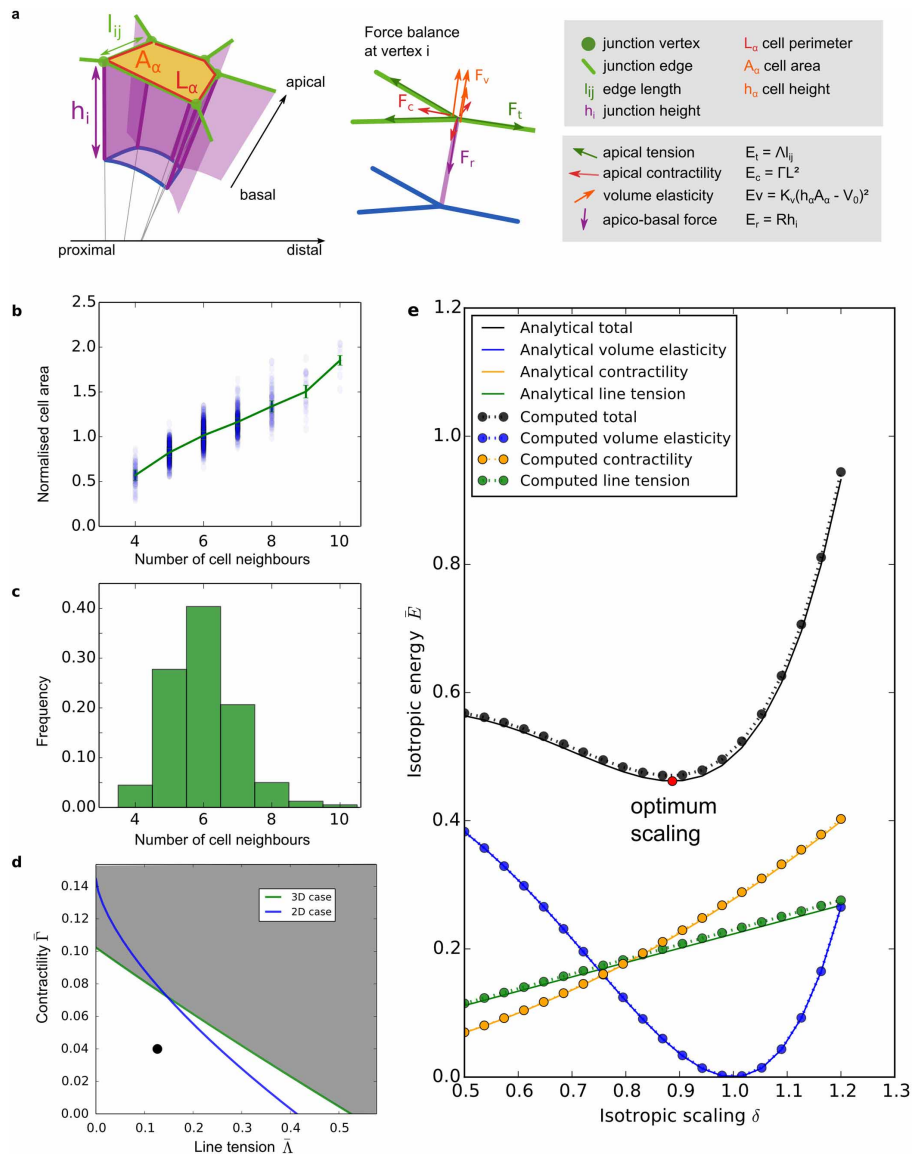
Extended Data Figure 3 | Myosin II and tension dynamics during fold formation. This figure is associated with Fig. 2. **a**, Close-up views of cells expressing MRLC::GFP, showing that, in the leg disc, myosin II is preferentially accumulated at the cortex (green) rather than at the medio-apical web (magenta). Relative intensity of myosin II at both levels is shown on separate panels ($n = 22$). **b**, Sagittal section of a leg disc from pre-fold stage (WP) showing that myosin II (visualized by MRLC::GFP construct, in green) co-localizes with E-Cad at adherens junctions (labelled in red) ($n = 22$). **c**, Quantification of myosin II::GFP (left) and F-actin (right) levels at individual junctions in fold (t4-t5) and segment (t4) domains of leg discs from pre-fold (young WP + 1 h) and late-fold (young WP + 4 h) stages ($n = 144$ and $n = 112$, respectively). Values are represented as mean values with error bars representing standard errors. The intensity of the signal in the fold has been normalized with the mean intensity in the segment domain. We used the non-parametric Wilcoxon rank sum test (also called Mann and Whitney test). **d, e**, Quantification of myosin II::GFP (left) and F-actin (right) levels per surface unit in fold (t4-t5) and segment (t4) domains of (d) leg discs from pre-fold (young WP + 1 h) and late-fold (young WP + 4 h) stages ($n = 23$ and $n = 18$ measurements respectively) or (e) in dissected discs cultured from pre-fold to mid-fold stage with either DMSO (control) or Q-VD-OPh (cell death inhibition) ($n = 32$ and $n = 24$ measurements respectively). Values are represented as mean values with error bars representing standard errors. The intensity of the signal in the fold has been normalized with the mean intensity in the segment domain. We used the non-parametric Wilcoxon rank sum test (also called Mann and Whitney test). **f**, Sagittal sections of pupal leg discs from mid-fold stage (WP + 2 h) of the following genotypes: *Dll-Gal4^[MD23]* (ctl, $n = 8$), *UAS-DIAP1;LP30-Gal4* (*Diap1*, $n = 27$), *Dll-Gal4^[MD23] / UAS-p35* (*p35*, $n = 11$) and *Dll-Gal4^[MD23] / UAS-p35; UAS-p35* (*p35 x2*, $n = 15$) (at early pupal stages, LP30-Gal4 and *Dll::Gal4^[MD23]* show

similar expression patterns, namely expression in the distal tibia and in all tarsal segments²⁷). The stabilization of myosin II and F-actin in the t4-t5 fold observed in the control (red arrowheads) is reduced or absent when cell death is inhibited (open arrowheads point out to the t4-t5 domain in the context of cell death inhibition). Myosin II is detected using anti-Sqh/MRLC antibody. The fold domain is false-coloured in pink and the segment domain in yellow. **g**, Quantification of myosin II (using anti-sqh antibody, left) and F-actin (right) levels per surface unit in fold (t4-t5) and segment (t4) domains in leg discs from mid-fold stage (WP + 2 h) in control (*DllGal4*) and cell inhibition (*Dll>p35x2*) contexts ($n = 25$ and $n = 28$ measurements respectively). Values are represented as mean values with error bars representing standard errors. The intensity of the signal in the fold has been normalized with the mean intensity in the segment domain. We used the non-parametric Wilcoxon rank sum test (also called Mann and Whitney test). **h**, Laser ablation experiments of apical membranes in *arm::GFP* leg discs in the segment domain versus the fold domain where apoptosis takes place. Discs were dissected and cultured *ex vivo* from pre-fold stage (WP) to mid-fold stage. Note the increase in the length of vertex release in the fold domain compared to the segment domain ($n = 4$). **i**, Laser ablation experiments of apical membranes in the fold domain of *arm::GFP* leg discs incubated from pre-fold stage (WP) to mid-fold stage with either DMSO (control, $n = 5$) or Q-VD-OPH (cell death inhibition, $n = 4$). Note that vertex release in the fold is reduced in the absence of apoptosis. **h, i**, Right panels, graphs representing quantifications of the increase in distance between vertices following laser cut, revealing apoptosis-dependent increased cellular tension in the fold domain versus the segment domain. Examples of ablated cells before (green) and after (magenta) laser cut are shown on the left. Orange bars represent the region where the laser cut has been performed. Errors bars correspond to the standard error of the mean.



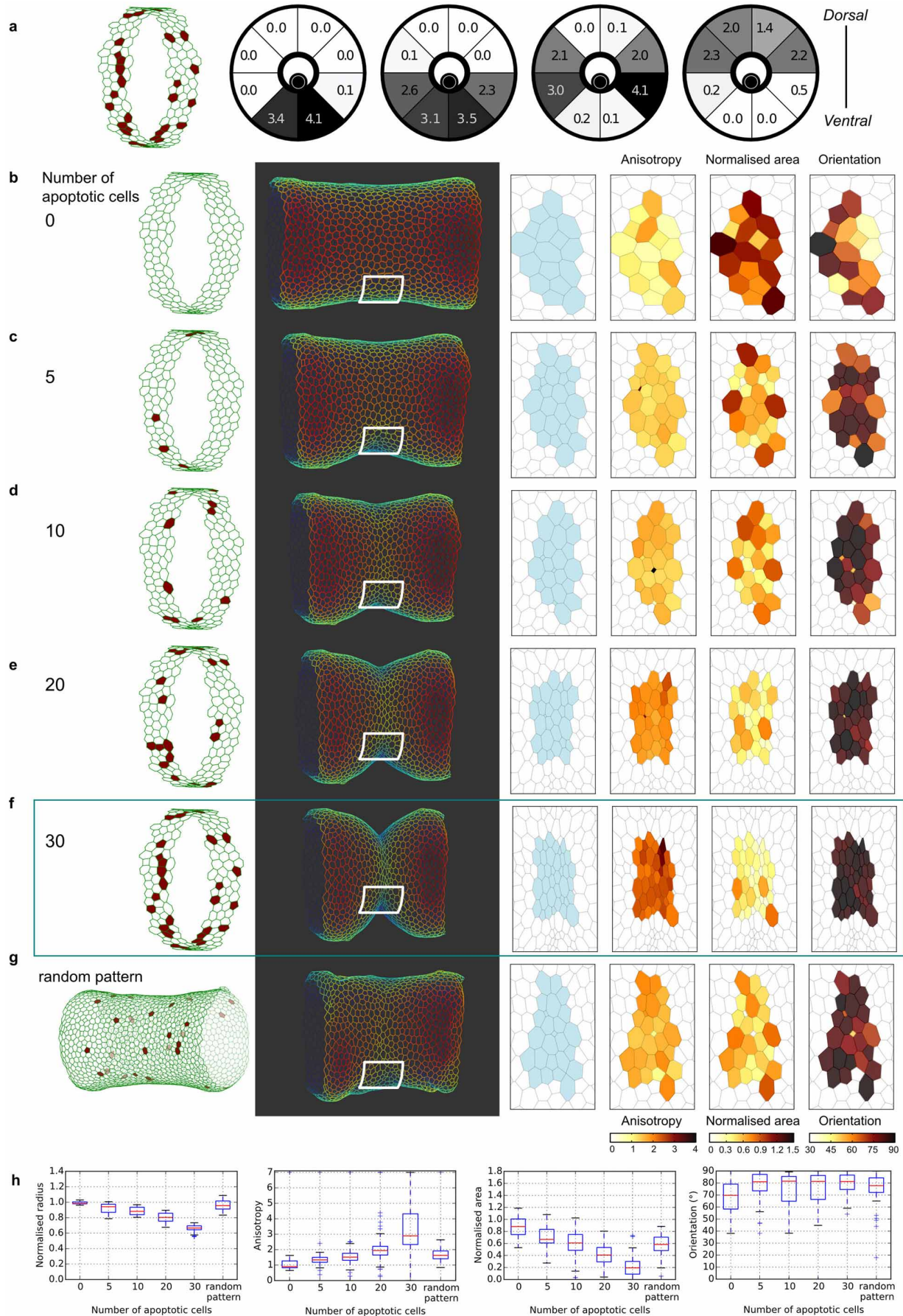
Extended Data Figure 4 | Cell shape dynamics during fold formation. This figure is associated with Fig. 2. **a**, Three-dimensional reconstructions and schematics of leg imaginal discs from pre-fold stage (WP) and mid-fold stage (WP + 2 h) stained with E-Cad in white and cleaved Dcp1 (revealing caspase activity) in red ($n = 11$ for each). High magnifications of the fold domain (surrounded in red) are shown on the right hand side of each panel. Arrowheads indicate apoptotic cells. **b**, Cell shape dynamics during fold formation ($n = 8$). **c**, Schematic of a pupal leg disc showing the *apterous* domain (in red) in the t4 tarsal segment and overlapping the t4-t5 fold (top). The bottom panel shows a typical result of automated cell outline extraction of a pupal leg disc double-stained for adherens junctions and the *apterous* domain. Cells from this domain have been subdivided into 12 sections along the proximo-distal axis (see Methods). **d**, Anisotropy, area and orientation of cells from DMSO (control, $n = 7$) or Q-VD-OPH (cell death inhibition, $n = 8$) leg discs from fold domain sections 10 and 11 (see Extended Data Fig. 4c) were

quantified and values represented as box plot. Discs were dissected and incubated from pre-fold stage (WP) to mid-fold stage. We used the non-parametric Wilcoxon rank sum test (also called Mann and Whitney test). **e**, Three-dimensional reconstruction images of anti E-Cad stained pupal leg discs from mid-fold stage (WP + 2 h) (top left) of the following genotypes: LP30::Gal4 (control, *ctl*, $n = 11$), *uas-DIAP1*; LP30::Gal4 (*Diap1*, $n = 22$), *Dll::Gal4[MD23]* / *uas-p35* (*p35*, $n = 9$) and *Dll::Gal4[MD23]* / *uas-p35*; *uas-p35* (*p35 x2*, $n = 8$) (at early pupal stages, LP30-Gal4 and *Dll::Gal4[MD23]* show similar expression patterns, namely expression in the distal tibia and in all tarsal segments²⁷). For each condition, cell outlines were extracted and anisotropy, apical surface area and orientation of cells from the fold domain were quantified and colour-coded. Note that when cell death is inhibited, anisotropy is reduced, apical surfaces are increased and cell preferential orientation is perturbed compared to the control situation.



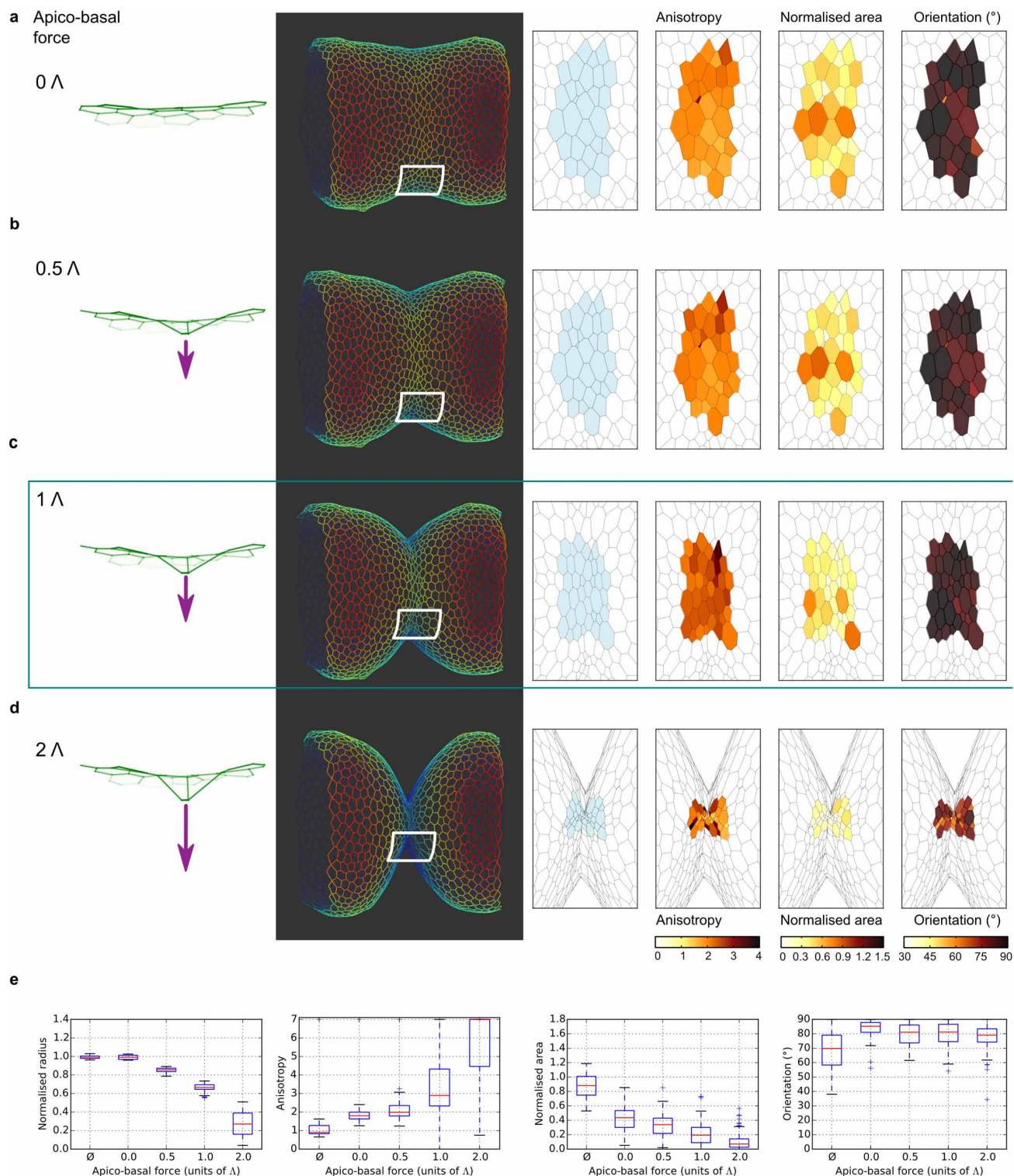
Extended Data Figure 5 | Three-dimensional epithelial cell model. This figure is associated with Fig. 3. **a**, Each cell is represented as an apical surface delimited by apical junctions. Each cell interacts with its neighbours through the apical junctions at its borders. In the original work by Farhadifar *et al.*¹⁵, three interactions are considered: (1) the tension opposing the elongation of a particular junction edge, with energy increasing with edge length; (2) a contractility, with energy proportional to the cell perimeter squared, used to model cell constriction; (3) a surface elasticity bringing the apical cell area back to a preferred area. As in the original work, the model can produce cell division, types one and three transitions, to which we added apoptosis. Yet in our case, contrary to Jülicher and colleagues work¹⁵, we must also take into account non planar modifications of the epithelial sheet. To this end, we modified the elastic area interaction to take into account a constrain on cell volume. The new interaction is termed volume elasticity and transmits contractions and dilations of the apical sheet along the apical-basal axis. The associated energy is proportional to the square of the difference between the current cell volume and a preferred volume. **b**, Average areas of cells as a function of the number of neighbouring cells in the epithelium before apoptosis, to be compared with Fig. 2g in Farhadifar *et al.*¹⁵. Our tissue shows a

similar trend of growing area with the number of sides, in good quantitative agreement with Farhadifar *et al.*¹⁵. **c**, Distribution of the number of neighbouring cells (or, equivalently, of the number of cell sides), to be compared with Fig. 2f of Farhadifar *et al.*¹⁵; once again, we are in good quantitative agreement with their model. **d**, Ground state diagram of the vertex model, comparing two-dimensional and three-dimensional hexagonal network boundaries (we restricted ourselves to $\Lambda > 0$ and $\Gamma > 0$ regions). The black dot indicates the chosen values for the line tension and contractility parameters, which are the same as case I in the Farhadifar *et al.*¹⁵ article. **e**, Variation of the normalized energy of a regular epithelium comprised of identical hexagonal cells as a function of a scale factor δ . Plain lines, analytical calculus; dotted line, average cell energy for a cylindrical tissue of 32 cells in diameter per 29 cells long. A scaling of $\delta = 1$ means that the cells are at their equilibrium volume in the absence of elasticity and contractility, and thus corresponds to the minimum of the blue lines (volume elasticity). Green lines correspond to line tension and yellow lines to contractility. The discrepancy between theoretical and computed values is due to the effect of cells lying at the border of the cylindrical epithelium.



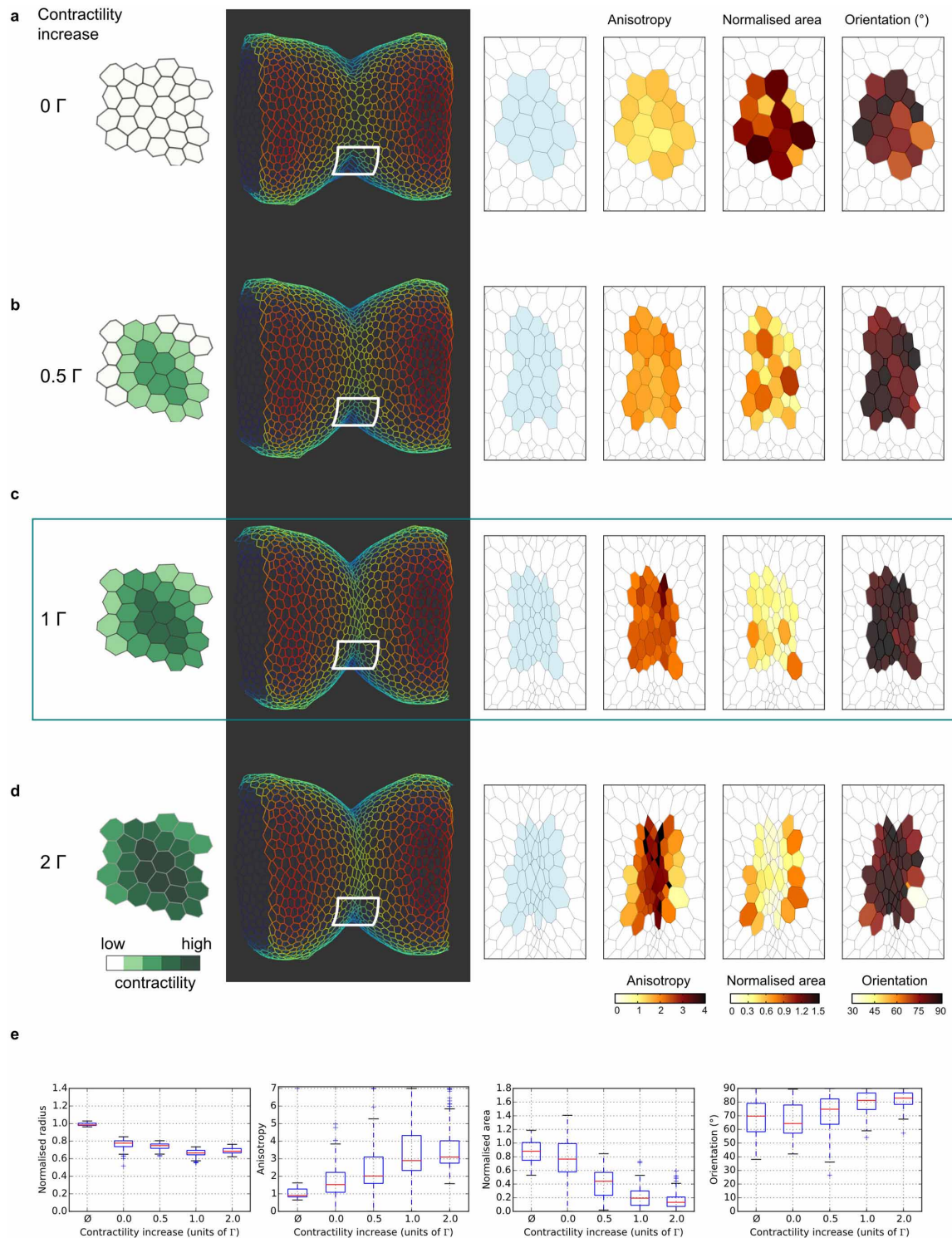
Extended Data Figure 6 | Effect of the apoptosis pattern on epithelium folding *in silico*. This figure is associated with Fig. 3. **a**, Representation of the *in silico* cell death pattern. Note the similarity with the *in vivo* distribution of apoptosis observed in biological samples (Fig. 1b). Nonetheless, the representation is not exactly comparable since cell death pattern is represented relative to a developmental stage in the biological samples, whereas in the model, the cell death pattern is represented relatively to the number of dead cells generated by the theoretical simulation since the time scale is not taken into account in the model. Left, representation of the *in silico* distribution of apoptotic cells around the fold domain. **b–g**, For each panel, from left to right are represented (1) a scheme of the fold domain showing the pattern of apoptosis, (2) a three-dimensional representation showing whole tissue shape and (3) for each condition, the corresponding cell outlines extracted from three-dimensional simulations in which anisotropy, area and orientation of cells from the fold domain are colour-coded. **b–f**, *In silico* models showing

whole tissue shape with an increasing number of apoptotic cells following the *in vivo* pattern of apoptosis. Note the gradual increase in anisotropy, gradual decrease of cell area and the preferential orientation with the gradual increase in the number of dying cells. The three-dimensional simulations in **b** and **f** are those presented in Fig. 3a and Fig. 3c, respectively. The three-dimensional simulation in **f** (framed in a blue rectangle in Extended Data Fig. 6, Extended Data Fig. 7 and Extended Data Fig. 8) corresponds to 30 apoptotic cells, with an apico-basal force of 1Λ and an increase in apical contractility of 1Γ . **g**, *In silico* model for a random pattern of apoptosis, with all other parameters similar to **f**. **h**, The mean value of radius, anisotropy, area and orientation of cells from the whole fold domain, defined as a $\pm 1 \mu\text{m}$ region around the fold centre of simulations **b–g** are represented by box plot. The number of cells considered is $n = 33, 38, 45, 61$ and 98 for $0, 5, 10, 20$ and 30 cells, respectively. This number varies due to changes in cells density in the domain.



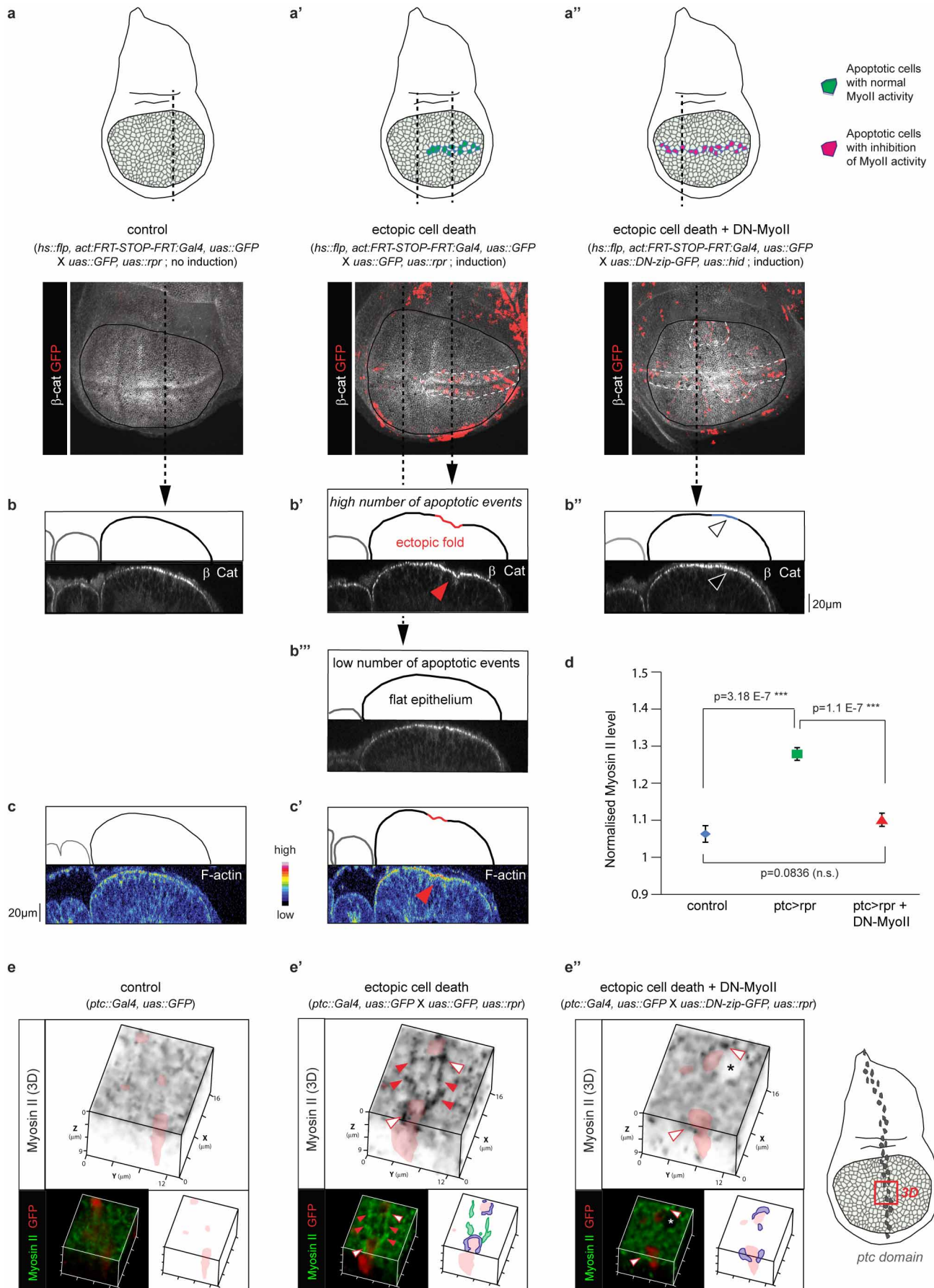
Extended Data Figure 7 | Effect of apico-basal force intensity on epithelium folding *in silico*. This figure is associated with Fig. 3. **a–d**, *In silico* models showing whole tissue shape following increasing values of apico-basal apoptotic force (30 apoptotic cells, apical contractility: 1 Γ). For each panel, from left to right are represented (1) a scheme of the strength of the apico-basal force applied, (2) a three-dimensional representation showing whole tissue shape and (3) for each condition, the corresponding cell outlines extracted from three-dimensional simulations in which anisotropy, area and orientation of

cells from the fold domain are colour-coded. Note the gradual increase in anisotropy, gradual decrease of cell area and the preferential orientation with the gradual increase of the apico-basal force. **e**, The mean value of radius, anisotropy, area and orientation of cells from the whole fold domain of simulations **a–d** are represented by box plot. The \emptyset symbol corresponds to the condition in absence of apoptosis (from Extended Data Fig. 6b). The number of cells considered is $n = 60, 69, 98$ and 157 for $\Lambda = 0, 0.5, 1$ and 2 , respectively.



Extended Data Figure 8 | Effect of the gradual apical cell contractility increase on epithelium folding *in silico*. This figure is associated with Fig. 3. **a–d**, *In silico* models showing whole tissue shape with an apico-basal force of 1Λ in 30 apoptotic cells and increasing values of apical contractility in neighbouring cells. For each panel, from left to right are represented (1) a scheme representing the gradual increase of contractility values applied in apoptotic neighbours, (2) a three-dimensional representation showing whole tissue shape and (3) for each condition, the corresponding cell outlines extracted from three-dimensional simulations in which cell anisotropy, area

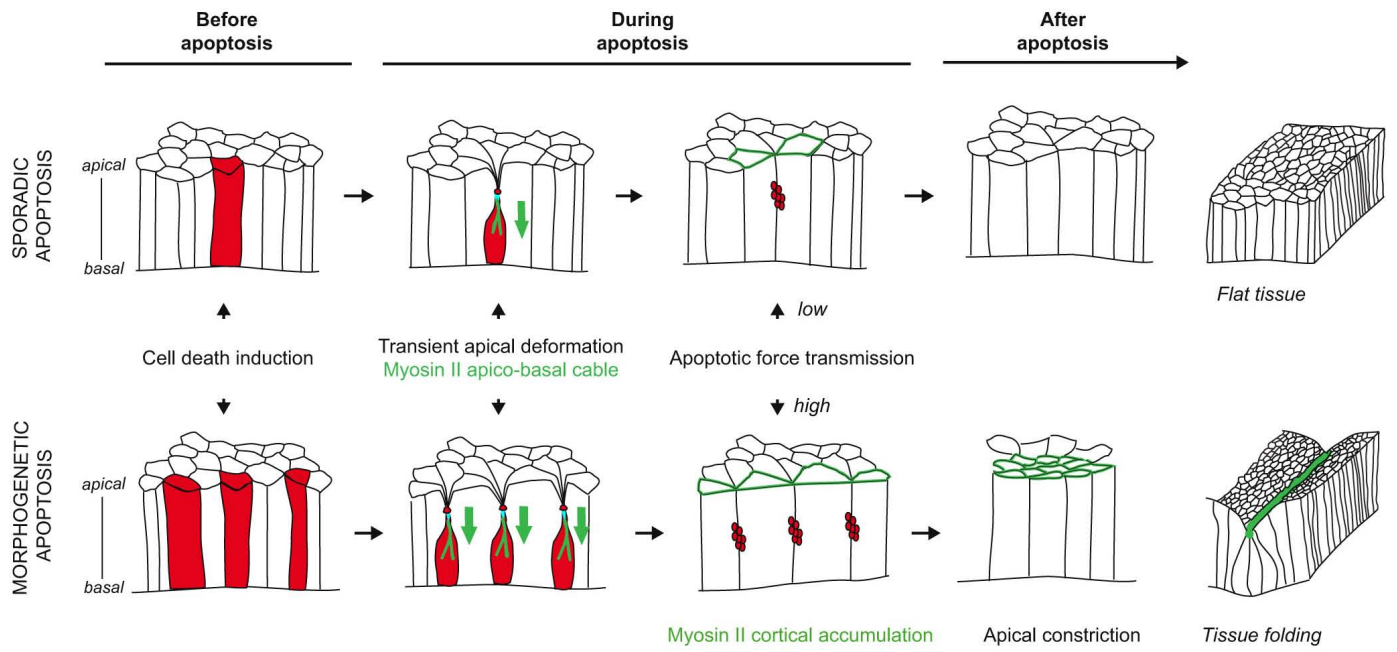
and orientation from the fold domain are colour-coded. Note the gradual increase in anisotropy, gradual decrease of cell area and the preferential orientation with the gradual increase of contractility. **e**, The mean value of radius, anisotropy, area and orientation of cells from the whole fold domain of simulations **a–d** are represented by box plot. The \emptyset symbol corresponds to the condition in absence of apoptosis (from Extended Data Fig. 6b). The number of cells considered is $n = 31, 61, 98$ and 137 for $\Gamma = 0, 0.5, 1$ and 2 , respectively.



Extended Data Figure 9 | Fold formation as a result of apoptotic forces.

This figure is associated with Fig. 4. **a–a''**, Schematics of wing discs depicting the pattern of apoptosis and the presence or absence of myosin II activity in apoptotic cells in each genetic context analysed in **b–b'''** and **c–c'**. Control (**a**, **b**, *y,w,hs::flp; act-*frt*-y+ -*frt*::Gal4, *uas::GFP / uas::lifeactGFP; uas::rpr* without clones), Ectopic cell death (**a'**, **b'**, **b'''**, *y,w,hs::flp; act-*frt*-y+ -*frt*::Gal4, *uas::GFP / uas::lifeactGFP; uas::rpr*), and ectopic cell death without myosin II activity specifically in apoptotic cells (**a''**, **b''**, *y,w,hs::flp; uas::DN-*zip*::GFP, *uas::hid; act-*frt*-CD2-*frt*::Gal4*). Note that, in the latter condition, myosin II activity is maintained in living cells. Wing discs were dissected from larvae heat shocked for 20 min at 38 °C. **b–c'**, For each panel, sagittal views and schematics of sagittal sections are shown (sagittal views correspond to the black dotted line indicated in **a–a''**). **b–b'''**, A high concentration of myosin II positive apoptotic cells is sufficient to induce a fold in a naive tissue (red arrowhead, **b'**, *n* = 11) as shown by the visualization of the wing disc apical surface stained with an anti- β -catenin antibody (compare **b'** with **b**). Note that no ectopic fold is observed when only a low number of apoptotic events occur (**b'''**) or when myosin II is inhibited in apoptotic cells (open arrowhead, *n* = 5 out of 6) (**b''**). **c–c'**, F-actin accumulates in ectopic folds (red arrowhead) when apoptotic myosin II is active in ectopic dying cells (compare **c'** with **c**, *n* = 3). **d**, Quantification of myosin II levels in the *patch* (*ptc*) domain***

(domain of cell death induction) in control (*w; ptc::Gal4, uas::GFP; tub::G80[ts] / SM5-TM6B*); *ptc* > *rpr* (*w; ptc::Gal4, uas::GFP / uas::lifeactGFP; tub::G80[ts] / uas::rpr*) and *ptc* > *rpr*+DN-MyoII (*w; ptc::Gal4, uas::GFP / uas::DN-*zip*::GFP; tub::G80[ts] / uas::rpr*) wing discs (control, *n* = 24; *ptc* > *rpr*, *n* = 28; *ptc* > *rpr*+DN-MyoII, *n* = 32). Values are represented as mean values with error bars representing standard errors. The intensity of the signal in the *ptc* domain has been normalized with the intensity in the anterior and posterior domains of the same disc (n.s. is for non-significant). We used the non-parametric Wilcoxon rank sum test (also called Mann and Whitney test). **e–e''**, Wing disc close-ups (of wing discs shown in Fig. 4c) and schematics in the absence (**e**) or presence (**e'** and **e''**) of ectopic apoptosis in the *ptc* domain (red cells, false-coloured in red on the black and white images), with (**e'**) or without (**e''**) myosin II activity in dying cells. **e'**, Note that we can distinguish two distinct pools of stabilized apical myosin II: “contractile ring myosin II” required for dying cell extrusion¹⁶ (blue arrows, purple in schematics) and “fold domain apical myosin II” stabilized in response to the apico-basal apoptotic force (red arrows, green in schematics). **e''**, Note that, consistently with normal extrusion in this background, contractile ring myosin II is still present around apoptotic cells (blue arrows), whereas fold domain apical myosin II is absent. The star points at a dividing cell, further indicating that myosin II is still present in apoptotic cell neighbours.



Extended Data Figure 10 | Model of apoptosis-dependent epithelium folding.

RNA helicase DDX21 coordinates transcription and ribosomal RNA processing

Eliezer Calo^{1*}, Ryan A. Flynn^{2*}, Lance Martin², Robert C. Spitale², Howard Y. Chang² & Joanna Wysocka^{1,3}

DEAD-box RNA helicases are vital for the regulation of various aspects of the RNA life cycle¹, but the molecular underpinnings of their involvement, particularly in mammalian cells, remain poorly understood. Here we show that the DEAD-box RNA helicase DDX21 can sense the transcriptional status of both RNA polymerase (Pol) I and II to control multiple steps of ribosome biogenesis in human cells. We demonstrate that DDX21 widely associates with Pol I- and Pol II-transcribed genes and with diverse species of RNA, most prominently with non-coding RNAs involved in the formation of ribonucleoprotein complexes, including ribosomal RNA, small nucleolar RNAs (snoRNAs) and 7SK RNA. Although broad, these molecular interactions, both at the chromatin and RNA level, exhibit remarkable specificity for the regulation of ribosomal genes. In the nucleolus, DDX21 occupies the transcribed rDNA locus, directly contacts both rRNA and snoRNAs, and promotes rRNA transcription, processing and modification. In the nucleoplasm, DDX21 binds 7SK RNA and, as a component of the 7SK small nuclear ribonucleoprotein (snRNP) complex, is recruited to the promoters of Pol II-transcribed genes encoding ribosomal proteins and snoRNAs. Promoter-bound DDX21 facilitates the release of the positive transcription elongation factor b (P-TEFb) from the 7SK snRNP in a manner that is dependent on its helicase activity, thereby promoting transcription of its target genes. Our results uncover the multifaceted role of DDX21 in multiple steps of ribosome biogenesis, and provide evidence implicating a mammalian RNA helicase in RNA modification and Pol II elongation control.

RNA helicases are highly conserved enzymes that use the energy of ATP to remodel RNA secondary structures and ribonucleoprotein complexes^{2,3} during various steps of RNA metabolism. In particular, the nucleolar helicase DDX21 is required for pre-rRNA processing^{4,5}, but the specific mechanism underlying this requirement remains unknown. Notably, DDX21 also influences c-Jun⁶ transcriptional activities, suggesting a potential role in gene expression. To explore this, we first interrogated the chromatin association of DDX21 in HEK293 cells by chromatin immunoprecipitation followed by high-throughput DNA sequencing (ChIP-seq). Given that pre-rRNA processing occurs coordinately with rDNA transcription, we examined binding of DDX21 to the rDNA locus (Fig. 1a). DDX21 broadly, but specifically, associated with the transcribed region of the rDNA, but not with the intergenic spacer, a profile characteristic of known Pol I-associated co-transcriptional regulators^{7,8}. In addition to rDNA binding, we identified 4,420 high-confidence peaks, most residing within 5 kilobases (kb) from annotated Pol II transcriptional start sites (Fig. 1b). DDX21-bound promoters had, on average, high enrichment of Pol II and active chromatin marks (histone H3 Lys 4 trimethylation (H3K4me3), H3K27 acetylation (H3K27ac) and H3K9ac), but were depleted for repressive (H3K27me3 and H3K9me3) and promoter-distal (H3K4me1) marks (Fig. 1c, d). Analysis of transcription factor motifs enriched at DDX21-bound regions uncovered recognition motifs of factors implicated in cell growth and proliferation (for example, E2F, STAT1, NRF1 and ETS; Extended Data Fig. 1a). ChIP-seq results were

verified by ChIP-qPCR (quantitative PCR) in two additional human cell lines, with all interrogated target regions showing enrichment by qPCR (Extended Data Fig. 1b and data not shown), indicating that the chromatin interactions of DDX21 are reproducible across multiple cell types.

Gene Ontology analyses of DDX21-bound regions revealed specific and highly significant association with several regulatory arms of the ribosomal pathway (Fig. 1e). To verify this further, we compared annotations of DDX21-bound promoters to those H3K4me3-enriched but DDX21-unbound (Extended Data Fig. 1c). As expected, DDX21-bound promoters were enriched for ribosomal Gene Ontology terms, while DDX21-unbound promoters were enriched for other biological processes (Extended Data Fig. 1d). DDX21 binding was evident at promoters of genes encoding components of both the 40S (for example, *RPS3*) and 60S (*RPL23A* and *RPL8*) subunits (Fig. 1f). Messenger RNAs that encode ribosomal proteins often harbour snoRNAs in their introns⁹. DDX21 binds promoters of more than 80% of snoRNA-containing host genes; those unbound are poorly expressed in HEK293 cells (Fig. 1g and Extended Data Fig. 1e–g).

To examine the effect of DDX21 on transcription, we depleted the protein using two independent short interfering RNA (siRNA) pools (Fig. 1h and Extended Data Fig. 2a). *DDX21* knockdown decreased the steady-state levels of transcripts originating from DDX21-bound promoters, but had minimal effect on the unbound gene transcripts (Fig. 1h and Extended Data Fig. 2b). To explore whether DDX21 directly regulates transcription of ribosomal mRNAs, we measured the effect of DDX21 depletion on the synthesis of nascent transcripts upon release from the transcriptional elongation block induced by the kinase inhibitor flavopiridol^{10,11}. We transfected HEK293 cells with control or *DDX21* 3' untranslated region (UTR) siRNAs, followed by expression of siRNA-resistant wild-type (DDX21^{WT}) or ATPase-defective¹² DDX21 (herein DDX21^{SAT}; Fig. 1i and Extended Data Fig. 2c). *DDX21* knockdown impaired the production of nascent transcripts originating from DDX21-bound promoters, and this effect was rescued by the introduction of DDX21^{WT}, but not DDX21^{SAT} (Fig. 1j and Extended Data Fig. 2d). Similar results were obtained on transcripts originating from the rDNA locus (Extended Data Fig. 2e). By contrast, non-target genes were minimally affected by the loss or ectopic expression of DDX21 (Fig. 1j and Extended Data Fig. 2d). Thus, DDX21 associates with and positively regulates transcription of Pol I- and Pol II-dependent ribosomal genes in a helicase-dependent manner.

The aforementioned observations prompted us to investigate potential crosstalk between DDX21 functions across nuclear compartments. Consistent with previous studies¹³, inhibition of Pol I with either CX-5461 or a low dose of actinomycin-D (ref. 14) displaced DDX21 from the nucleolus, whereas localization of the nucleolar protein fibrillarin was not affected under these conditions (Fig. 2a and Extended Data Fig. 3a, b). Notably, hour-long inhibition of Pol II with flavopiridol recapitulated the nucleolar exclusion of DDX21 (Fig. 2a). By contrast, serum starvation or treatment with metabolic inhibitors impacting either cellular respiration or the mTOR pathway did not alter DDX21 localization

¹Department of Chemical and Systems Biology, Stanford University School of Medicine, Stanford, California 94305, USA. ²Howard Hughes Medical Institute and Program in Epithelial Biology, Stanford University School of Medicine, Stanford, California 94305, USA. ³Department of Developmental Biology, Stanford University School of Medicine, Stanford, California 94305, USA.

*These authors contributed equally to this work.

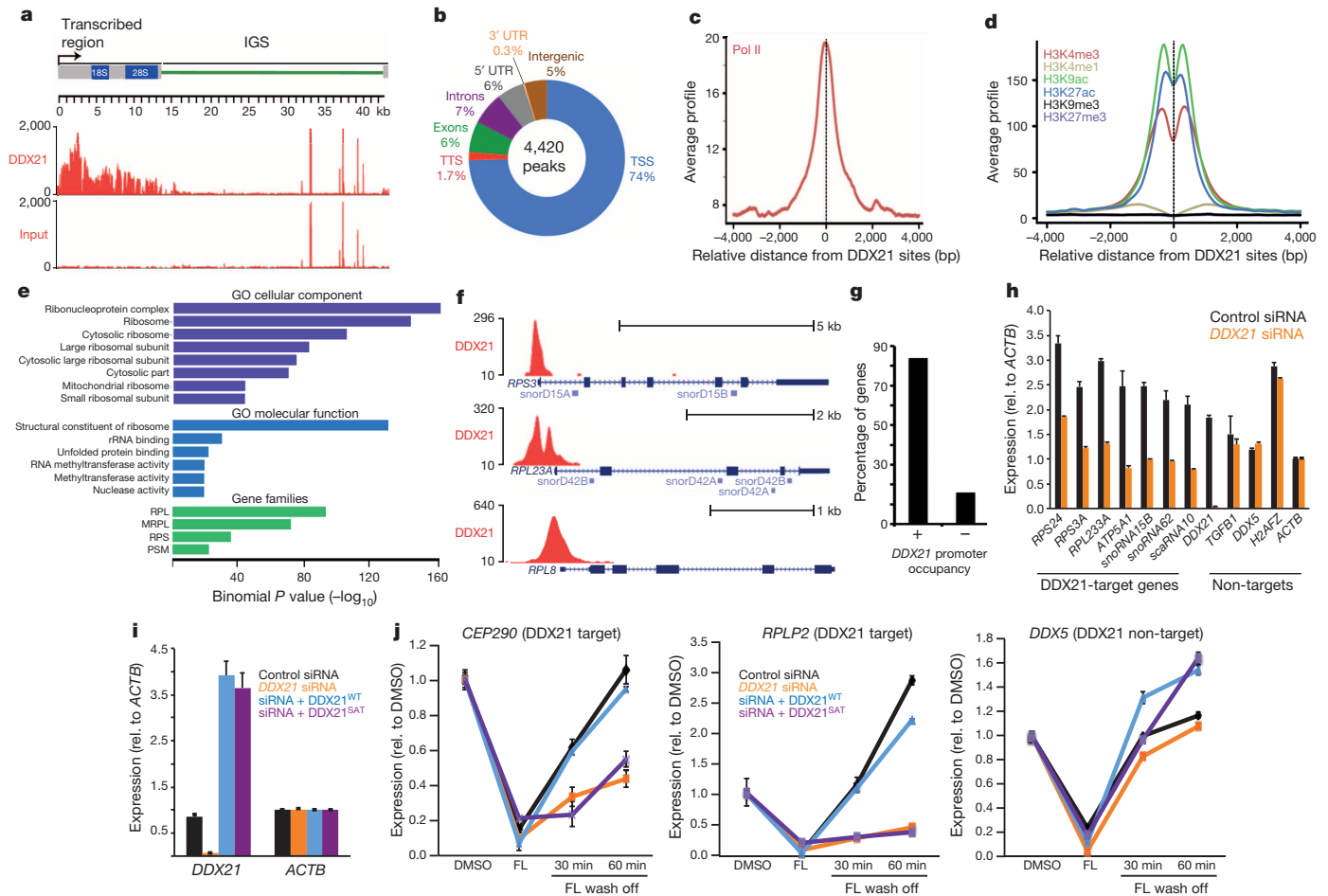


Figure 1 | DDX21 associates with actively transcribed ribosomal genes. **a**, DDX21 ChIP-seq reads were mapped to a custom annotation file of the human rDNA locus and compared to input reads. IGS, intergenic spacer. **b**, Distribution of DDX21 ChIP-seq peaks over known genomic features. TSS, transcription start site; TTS, transcription termination site. **c**, **d**, Average ChIP-seq signal profiles from publicly available data sets (see Methods) were generated for Pol II (**c**) and the indicated histone modifications around the centre of DDX21-bound regions (**d**). bp, base pairs. **e**, Genomic regions enrichment of annotations tool (GREAT) analyses of DDX21-bound regions. The x axis corresponds to the negative binomial P values. GO, Gene Ontology. **f**, University of California Santa Cruz (UCSC) genome browser tracks of DDX21 ChIP-seq at ribosomal genes containing intronic snoRNAs.

(Extended Data Fig. 4), underscoring the preferential sensitivity of DDX21 to transcriptional inhibition. Furthermore, inhibition of either Pol I or II impaired the association of DDX21 with both the rDNA and Pol II-regulated promoters (Fig. 2c and Extended Data Fig. 3c, d). This change in chromatin association was not due to widespread chromatin silencing and compaction, as inhibitors did not affect CTCF binding at the rDNA- or Pol II-regulated chromatin (Extended Data Fig. 3e, f). Therefore, the chromatin association of DDX21 relies on the transcriptional status of either Pol I or II, suggesting coordination of the functions of DDX21 across subnuclear compartments.

The roles of DDX21 in transcription and rRNA processing are dependent on its intact helicase domain^{4,6}. We proposed that defining the RNA interactome of DDX21 would reveal insights into the molecular mechanisms underlying its diverse functions. To identify DDX21-associated RNAs systematically, we performed tandem purification iCLIP (individual-nucleotide-resolution crosslinking and immunoprecipitation)¹⁵ (Fig. 3a and Extended Data Fig. 5a, b) in HEK293 cells induced to express Flag- and haemagglutinin-tagged DDX21 (Flag-HA-DDX21). DDX21 interacts with a diverse set of RNAs, of which rRNA and snoRNAs were most highly represented, while mRNAs contributed only 1.1% of

the iCLIP reads (Fig. 3b). Gene ontology term and KEGG pathway analysis linked these mRNAs to ribosome function (Fig. 3b). Comparisons between DDX21 iCLIP targets and those of the splicing factor hnRNP-C¹⁶ revealed little overlap, underscoring the specificity of our results (Extended Data Fig. 5c–e). We further confirmed select iCLIP interactions by ultraviolet RNA immunoprecipitation and quantitative reverse transcription PCR (qRT-PCR) (Extended Data Fig. 5f).

rRNA and snoRNAs represent candidate direct partners for DDX21-mediated rRNA processing function. DDX21 broadly crosslinks to rRNA, with the strongest binding overlapping 2'-O-methylation (2'-Ome) and pseudouridylation (Ψ) sites (Fig. 3c), which are targeted and modified by snoRNP complexes¹⁷. Furthermore, DDX21 robustly interacts with regions in the 5' external transcribed spacer, which is bound and processed by the U3 snoRNA¹⁸. Consistently, U3 is the most enriched short repetitive RNA of DDX21, crosslinking to DDX21 in two distinct 5' and 3' regions of known rRNA targeting function¹⁸ (Extended Data Figs 6f and 8c). iCLIP of DDX21^{SAT} revealed that it retained the ability to bind the major classes of RNA recovered with the wild-type enzyme (Extended Data Fig. 6a–d). However, there was a marked restriction of DDX21^{SAT} to 18S rRNA (Extended Data Fig. 6e). Notably, the inability

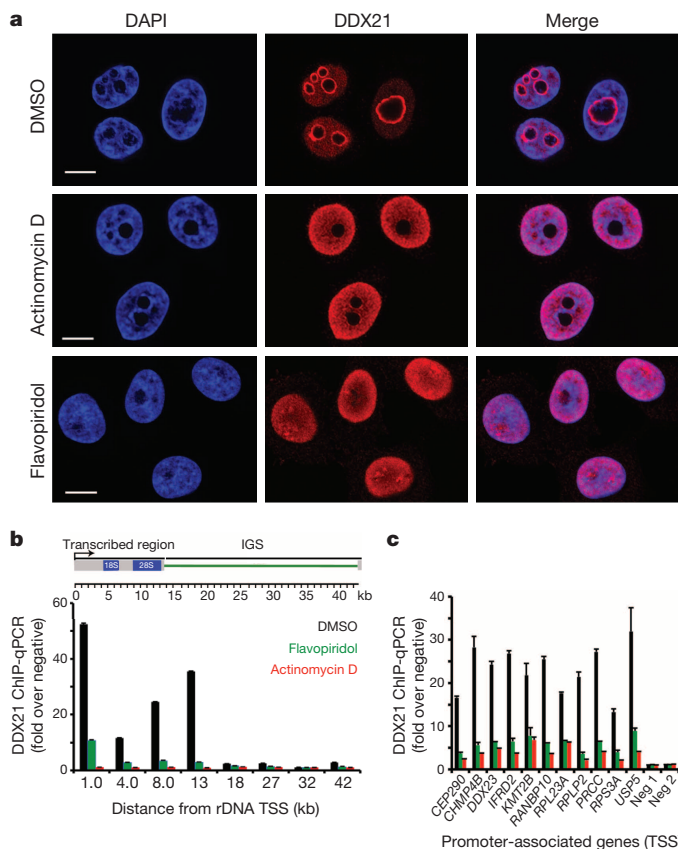


Figure 2 | DDX21 chromatin association is sensitive to Pol I and Pol II transcriptional status. **a**, Representative immunofluorescence images of methanol-fixed HEK293 cells after 1-h incubation with DMSO, actinomycin-D (50 ng ml⁻¹) or flavopiridol (1 μM). Scale bars, 10 μm. DAPI, 4',6-diamidino-2-phenylindole. **b**, **c**, ChIP-qPCR analysis in HEK293 cells sampling DDX21 genomic occupancy at the rDNA locus and at a representative panel of Pol I-regulated, DDX21-target promoters, upon treatment with DMSO, actinomycin-D or flavopiridol. Data are mean and s.d. of three independent experiments. Neg, negative controls.

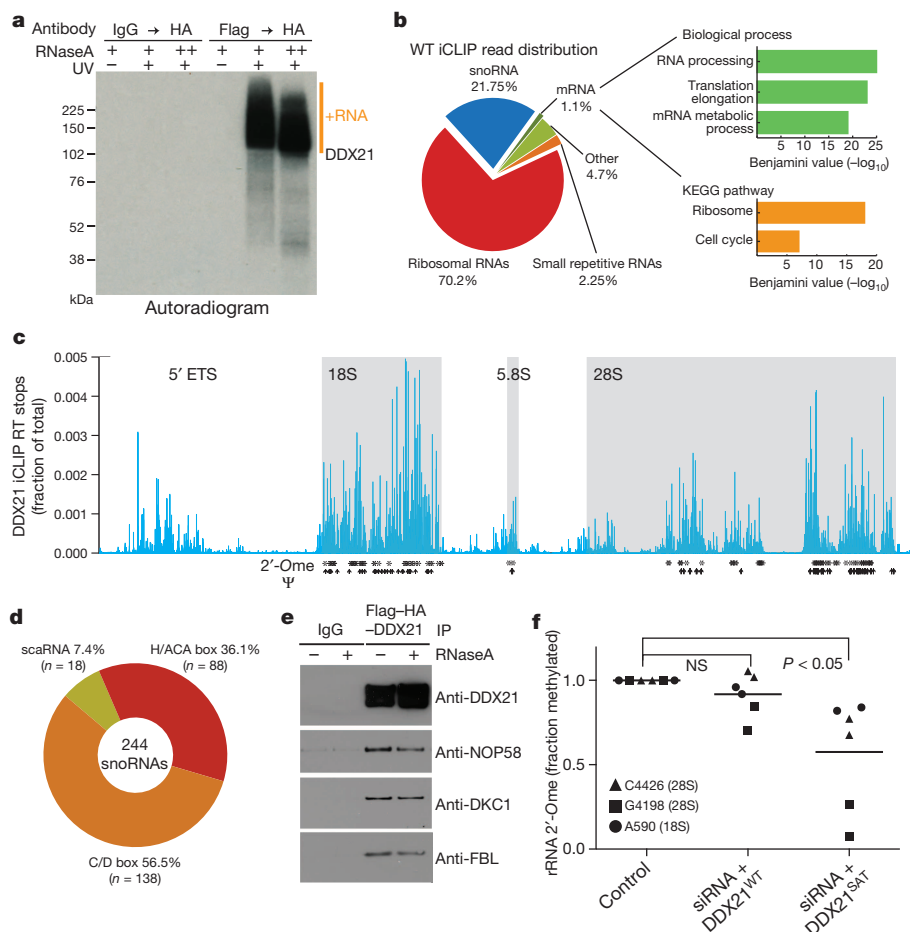
of DDX21^{SAT} to bind the 5' external transcribed spacer was accompanied by the loss of the 5' end binding to U3 (Extended Data Fig. 6f). Importantly, marked loss of DDX21^{SAT} association with the rRNA occurred in the absence of transcriptional defects in snoRNA production (Extended Data Fig. 6g).

Consistent with preferential association of DDX21^{WT} at rRNA modification sites, snoRNAs represent the principal class of DDX21-bound RNAs (Fig. 3d), as exemplified by snorD66 and snorA67, representing the C/D- and H/ACA-box subfamilies, respectively. DDX21 crosslinked most robustly outside the 'box' motifs (Extended Data Fig. 7a, b), a pattern distinct from canonical snoRNP-binding proteins^{19,20}. In addition to snoRNAs, protein components of the snoRNP also associate with DDX21. We recovered NOP58, fibrillarin and dyskerin in DDX21 immunoprecipitation, both in the presence and absence of RNaseA (Fig. 3e). Conversely, NOP58 co-immunoprecipitated DDX21 (Extended Data Fig. 7c). Of note, DDX21 also binds XRN2, a snoRNA-processing exonuclease²¹, but in an RNase-dependent manner (Extended Data Fig. 7d).

To explore whether DDX21 is required for snoRNP-mediated rRNA modifications¹⁷, we assayed changes in 2'-Ome levels using site-directed cleavage of rRNA by RNaseH²² (Extended Data Fig. 7e). We treated HEK293 cells with control or DDX21 3' UTR siRNAs, followed by expression of DDX21^{WT} or DDX21^{SAT}. Reintroduction of DDX21^{SAT}

Figure 3 | DDX21 binds rRNA and snoRNAs and facilitates rRNA modification.

a, Autoradiogram of ³²P-labelled RNA crosslinked to Flag-HA-DDX21^{WT}. RNA-protein complexes are seen in the purifications from Flag-HA-DDX21^{WT} cells but not from control cells. kDa, kilodaltons; UV, ultraviolet crosslinking. **b**, Left, DDX21 iCLIP reads annotated to known repetitive (rRNA and snRNAs) and non-repetitive regions of the human genome with percentage of the total iCLIP reads shown. Right, enriched Gene Ontology and KEGG pathway terms from DDX21-bound mRNAs obtained using the DAVID tool. The x axis values correspond to the negative Benjamini P value. **c**, DDX21^{WT} iCLIP reads mapped to the transcribed region of the rDNA. The mature portions of the 18S, 5.8S and 28S rRNAs are highlighted (grey). Known 2'-Ome (*) and Ψ (†) modification sites are marked below the iCLIP reads. RT, reverse transcription. ETS, external transcribed spacer. **d**, Distribution of all DDX21^{WT}-bound snoRNAs, representing C/D box, H/ACA box, and scaRNAs. The number (n) and fraction of each snoRNA type are displayed. **e**, Co-immunoprecipitation analysis of Flag-HA-DDX21^{WT} and components of the snoRNPs. Conditions with/without RNaseA treatment are shown. **f**, Site-directed cleavage of rRNA by RNaseH at three sites. Values from three biological replicates are shown. Student's t-test was performed and only DDX21^{SAT} was significantly reduced (P < 0.05) relative to control. NS, not significant.



failed to rescue rRNA methylation levels at multiple sites, especially on the 28S rRNA, whereas DDX21^{WT} recovered the 2'-Ome defects (Fig. 3f). We cannot exclude the possibility that the observed 2'-Ome defects were due to the role of DDX21 in snoRNA transcription. However, changes in snoRNA binding pattern between DDX21^{SAT} and DDX21^{WT} can be readily detected in the absence of snoRNA transcriptional defects (Extended Data Figs 6f, g and 7f). Therefore, the helicase domain of DDX21 is required for proper binding to both rRNA and snoRNAs, suggesting a direct function within the snoRNP.

We next examined our iCLIP results for insights into the role of DDX21 in regulating Pol II-dependent transcription. While ultraviolet crosslinking to mRNAs suggested a potential *cis*-mechanism at DDX21-target genes, transcripts from most DDX21-occupied promoters were not recovered in iCLIP and only 8% of the iCLIP mRNA reads mapped within 5' UTRs (Extended Data Fig. 8a, b). Thus, nascent RNA tethering is unlikely to be the major mechanism of DDX21 chromatin recruitment. Notably, 7SK snRNA, a well-known *trans*-acting non-coding RNA involved in Pol II transcription^{23–26}, was among the most highly enriched DDX21-bound RNAs (Extended Data Fig. 8c). Together with HEXIM1/2 and P-TEFb (which consists of a CDK9 and cyclin T1 heterodimer), 7SK functions to modulate Pol II promoter pause–release^{27,28}. We observed DDX21 crosslinking most robustly to two specific sites on 7SK (Fig. 4a), outside the known binding sites of HEXIM1/2 and P-TEFb²⁹ (Extended Data Fig. 8d). Recent evidence demonstrated that the 7SK snRNP is physically associated with Pol II promoters, where P-TEFb is then released^{24,25,30}. We proposed that DDX21 is recruited to Pol II promoters together with the 7SK snRNP. Consistent with this hypothesis, protein components of the 7SK snRNP associate with DDX21 in reciprocal co-immunoprecipitation experiments (Extended Data Fig. 8e, f). Furthermore, both HEXIM1 and CDK9 are bound at DDX21-target Pol II promoters, but not at the rDNA locus (Extended Data Fig. 9a, b). Finally, depletion of

7SK with two independent antisense oligonucleotides²⁶ strongly reduced the association of DDX21 with Pol II-bound promoters, but not at the rDNA (Fig. 4b and Extended Data Fig. 9c), indicating that 7SK facilitates the association of DDX21 with promoters. Notably, similar to DDX21, Gene Ontology analyses of CDK9 ChIP-seq from HEK293 cells³⁰ revealed enrichment for components of the ribosomal pathway (Extended Data Fig. 9d), suggesting that the 7SK snRNP complex might have an inherent preference for binding at ribosomal and growth control genes. Nonetheless, this preference alone cannot entirely explain the specificity of DDX21 targeting to ribosomal genes, because we also detected HEXIM1 and CDK9 at active promoters unbound by DDX21 (Extended Data Fig. 9a, b).

Transcriptional elongation is triggered by phosphorylation of the Pol II carboxy-terminal domain at serine 2 (Ser2p) upon release of P-TEFb from 7SK^{25,28,30}. To test whether DDX21 facilitates P-TEFb release, we performed a 'release assay'^{24,30} by purifying the inactive form of the 7SK snRNP. Incubation of the inactive 7SK snRNP with purified Flag–HA–DDX21 resulted in the dose-dependent release of CDK9, whereas a control Flag purification showed no release (Fig. 4c, d). Consistent with its role in promoting the activity of P-TEFb, DDX21 knockdown impaired Pol II Ser2p at the 3' ends of DDX21-target genes (Fig. 4e). Total levels of Pol II at the same regions were also diminished, concordant with an elongation defect (Extended Data Fig. 9e).

DDX21 promotes transcription in a manner dependent on its catalytic domain (Fig. 1j and Extended Data Fig. 2d) and we proposed that this feature would extend to its role in P-TEFb release. Notably, although DDX21^{SAT} is recruited to Pol II promoters, binds to 7SK and interacts with P-TEFb in the lysate (Extended Data Fig. 10a–c), it shows marked differences in the 7SK ultraviolet crosslinking as compared to DDX21^{WT}, with accumulation at the single-stranded region upstream from the fourth stem loop, involved in P-TEFb binding and inhibition²⁹ (Extended Data

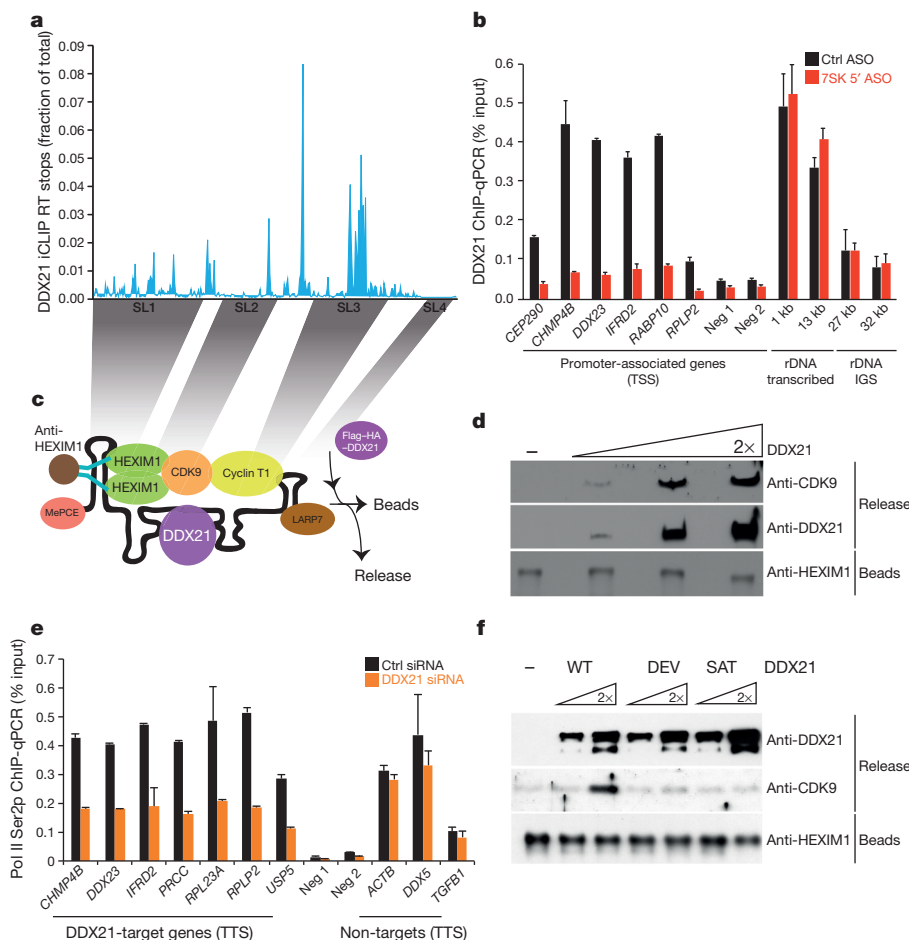


Fig. 10c). These observations suggest that DDX21^{SAT} recognizes 7SK, but is unable to remodel the RNA. Indeed, we find that both DDX21^{SAT} and an additional catalytically defective mutant, DDX21^{DEV} (Extended Data Fig. 2c), failed to release P-TEFb from the inactive 7SK snRNP (Fig. 4f). Therefore, DDX21 requires its active helicase domain to drive Pol II-dependent transcription through release of P-TEFb and consequently promote transcriptional elongation.

Collectively, we showed that DDX21 is incorporated into distinct snRNP complexes, 7SK snRNP and snoRNP (Extended Data Fig. 10d), to regulate transcriptional and post-transcriptional steps of ribosome biogenesis. Our data suggest that DDX21 is a key component in coordinating transcriptional programs across distinct nuclear compartments, as its engagement with chromatin is sensitive to the status of both Pol I and Pol II. We propose that through its multifaceted function in ribosome biogenesis, DDX21 has a key role in regulating cellular growth in health and malignancy.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 4 April; accepted 6 October 2014.

Published online 24 November 2014.

- Rocak, S. & Linder, P. DEAD-box proteins: the driving forces behind RNA metabolism. *Nature Rev. Mol. Cell Biol.* **5**, 232–241 (2004).
- Russell, R., Jarmoskaite, I. & Lambowitz, A. M. Toward a molecular understanding of RNA remodeling by DEAD-box proteins. *RNA Biol.* **10**, 44–55 (2013).
- Putnam, A. A. & Jankowsky, E. DEAD-box helicases as integrators of RNA, nucleotide and protein binding. *Biochim. Biophys. Acta* **1829**, 884–893 (2013).
- Henning, D., So, R. B., Jin, R., Lau, L. F. & Valdez, B. C. Silencing of RNA helicase II/Guα inhibits mammalian ribosomal RNA production. *J. Biol. Chem.* **278**, 52307–52314 (2003).
- Yang, H. *et al.* Down-regulation of RNA helicase II/Gu results in the depletion of 18 and 28 S rRNAs in *Xenopus* oocyte. *J. Biol. Chem.* **278**, 38847–38859 (2003).
- Westermarck, J. *et al.* The DEXD/H-box RNA helicase RHII/Gu is a co-factor for c-Jun-activated transcription. *EMBO J.* **21**, 451–460 (2002).
- Zentner, G. E., Saiakhova, A., Manaenkova, P., Adams, M. D. & Scacheri, P. C. Integrative genomic analysis of human ribosomal DNA. *Nucleic Acids Res.* **39**, 4949–4960 (2011).
- Cong, R. *et al.* Interaction of nucleolin with ribosomal RNA genes and its role in RNA polymerase I transcription. *Nucleic Acids Res.* **40**, 9441–9454 (2012).
- Dieci, G., Preti, M. & Montanini, B. Eukaryotic snoRNAs: a paradigm for gene expression flexibility. *Genomics* **94**, 83–88 (2009).
- Chao, S. H. *et al.* Flavopiridol inhibits P-TEFb and blocks HIV-1 replication. *J. Biol. Chem.* **275**, 28345–28348 (2000).
- Chao, S. H. & Price, D. H. Flavopiridol inactivates P-TEFb and blocks most RNA polymerase II transcription *in vivo*. *J. Biol. Chem.* **276**, 31793–31799 (2001).
- Valdez, B. C., Henning, D., Perumal, K. & Busch, H. RNA-unwinding and RNA-folding activities of RNA helicase II/Gu—two activities in separate domains of the same protein. *Eur. J. Biochem.* **250**, 800–807 (1997).
- Perlaky, L., Valdez, B. C. & Busch, H. Effects of cytotoxic drugs on translocation of nucleolar RNA helicase RH-II/Gu. *Exp. Cell Res.* **235**, 413–420 (1997).
- Drygin, D. *et al.* Anticancer activity of CX-3543: a direct inhibitor of rRNA biogenesis. *Cancer Res.* **69**, 7653–7661 (2009).
- Huppertz, I. *et al.* iCLIP: Protein-RNA interactions at nucleotide resolution. *Methods* **65**, 274–287 (2014).
- Zarnack, K. *et al.* Direct competition between hnRNP C and U2AF65 protects the transcriptome from the exonization of Alu elements. *Cell* **152**, 453–466 (2013).
- Lui, L. & Lowe, T. Small nucleolar RNAs and RNA-guided post-transcriptional modification. *Essays Biochem.* **54**, 53–77 (2013).
- Hughes, J. M. & Ares, M. Jr. Depletion of U3 small nucleolar RNA inhibits cleavage in the 5' external transcribed spacer of yeast pre-ribosomal RNA and impairs formation of 18S ribosomal RNA. *EMBO J.* **10**, 4231–4239 (1991).
- Fatica, A., Galardi, S., Altieri, F. & Bozzoni, I. Fibrillarin binds directly and specifically to U16 box C/D snoRNA. *RNA* **6**, 88–95 (2000).
- Newman, D. R., Kuhn, J. F., Shanab, G. M. & Maxwell, E. S. Box C/D snoRNA-associated proteins: two pairs of evolutionarily ancient proteins and possible links to replication and transcription. *RNA* **6**, 861–879 (2000).
- Petfalski, E., Dandekar, T., Henry, Y. & Tollervey, D. Processing of the precursors to small nucleolar RNAs and rRNAs requires common components. *Mol. Cell. Biol.* **18**, 1181–1189 (1998).
- Yu, Y. T., Shu, M. D. & Steitz, J. A. A new method for detecting sites of 2'-O-methylation in RNA molecules. *RNA* **3**, 324–331 (1997).
- Yang, Z., Zhu, Q., Luo, K. & Zhou, Q. The 7SK small nuclear RNA inhibits the CDK9/cyclin T1 kinase to control transcription. *Nature* **414**, 317–322 (2001).
- Ji, X. *et al.* SR proteins collaborate with 7SK and promoter-associated nascent RNA to release paused polymerase. *Cell* **153**, 855–868 (2013).
- McNamara, R. P., McCann, J. L., Gudipaty, S. A. & D'Orso, I. Transcription factors mediate the enzymatic disassembly of promoter-bound 7SK snRNP to locally recruit P-TEFb for transcription elongation. *Cell Rep.* **5**, 1256–1268 (2013).
- Castelo-Branco, G. *et al.* The non-coding snRNA 7SK controls transcriptional termination, pausing, and bidirectionality in embryonic stem cells. *Genome Biol.* **14**, R98 (2013).
- Nguyen, V. T., Kiss, T., Michels, A. A. & Bensaude, O. 7SK small nuclear RNA binds to and inhibits the activity of CDK9/cyclin T complexes. *Nature* **414**, 322–325 (2001).
- Peterlin, B. M. & Price, D. H. Controlling the elongation phase of transcription with P-TEFb. *Mol. Cell* **23**, 297–305 (2006).
- Michels, A. A. *et al.* Binding of the 7SK snRNA turns the HEXIM1 protein into a P-TEFb (CDK9/cyclin T) inhibitor. *EMBO J.* **23**, 2608–2619 (2004).
- Liu, W. *et al.* Brd4 and JMJD6-associated anti-pause enhancers in regulation of transcriptional pause release. *Cell* **155**, 1581–1595 (2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank D. H. Price for the LARP7 antibody, K. Lane from M. Covert's laboratory for metabolic inhibitors, K. Cimprich and members of the Chang and Wysocka laboratories for discussions, and B. Zarnegar and P. Khavari for discussions regarding iCLIP. This work was supported by the Stanford Medical Scientist Training Program and T32CA09302 (R.A.F.), AP Giannini Foundation (R.C.S.), National Institutes of Health grants R01-HG004361, R01-ES023168, P50-HG007735 (H.Y.C.) and R01-GM095555 (J.W.), W. M. Keck Foundation (J.W.), and Helen Hay Whitney Foundation (E.C.). H.Y.C. is an Early Career Scientist of the Howard Hughes Medical Institute.

Author Contributions H.Y.C. and J.W. supervised the project; E.C. and R.A.F. conceived and designed the study; E.C. and R.A.F. performed experiments and analysed ChIP-seq data. R.A.F. performed iCLIP and L.M., R.C.S. and R.A.F. analysed iCLIP data; R.A.F., E.C., J.W. and H.Y.C. wrote the manuscript with input from all co-authors.

Author Information All sequencing data have been deposited in Gene Expression Omnibus (GEO) data repository under accession number GSE56802. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.W. (wysocka@stanford.edu) or H.Y.C. (howchang@stanford.edu).

METHODS

HEK293 cells were cultured in DMEM plus 10% FBS and maintained under standard tissue culture conditions. ChIP-seq and ChIP-qPCR analyses were conducted in HEK293 cells with commercially available antibodies as described. The nascent transcript assays have been described elsewhere³¹. In brief, HEK293 cells stably expressing inducible transgenes for either DDX21^{WT} or DDX21^{SAT} were cultured in tetracycline-free serum (GIBCO) and seeded at a density of 30–50% confluency overnight. Cells were transfected for two consecutive days with a pool of siRNA targeting the *DDX21* 3' UTR (see later in the text). Twelve hours after the second transfection, 0.025 µg ml⁻¹ doxycycline (dox) was added to the media for an additional 24 h. At this point cells were treated for 1 h with either DMSO or flavopiridol (1.0 µg ml⁻¹) to inhibit Pol II elongation. For 'wash-off' cultures, cells were washed 3–4 times with pre-warmed media and allowed to recover transcription for either 30 or 60 min. Cells were immediately collected in TRIzol and RNA was extracted as described later. For immunofluorescence studies cells were fixed in methanol and stained for the indicated antibodies. Small molecule treatments with DMSO, actinomycin-D, flavopiridol or CX-5461 were done for 1 h, unless otherwise specified. iCLIP was performed from HEK293 cells stably expressing Flag-HA-DDX21. The RNaseH cleavage assay was optimized from a previously reported method to cleave RNA at specific bases in a manner sensitive to the 2'-Ome status of the nucleotide of interest²². For the P-TEFb release assay, the inactive 7SK snRNP complex was purified by immobilizing HEXIM1, followed by incubation with biochemically purified Flag-HA-DDX21. The predicted 7SK structure has been published elsewhere³².

Cell lines. HEK293 and HeLa cells were obtained from ATCC and grown under standard conditions in DMEM plus 10% FBS supplemented with antibiotics. All cell lines used in this study are mycoplasma-free. For generating stable HEK293 cells expressing Flag-HA-DDX21 (FH-DDX21) full-length *DDX21* was amplified from complementary DNA and cloned into a dox-inducible pTrip-Flag-HA lentiviral vector. The cloning strategy for generating DDX21^{DEV} and DDX21^{SAT} has been previously described¹². HEK293 cells were infected and several clones were expanded for further analyses and tested for mycoplasma. For the iCLIP experiments, expression of the *DDX21* transgenes was achieved by addition of dox (0.025 µg ml⁻¹) to the media. Cells were collected 24 h after the addition of dox.

ChIP-qPCR and ChIP-seq. ChIP assays were performed as previously described^{24,33}. In brief, HEK293 cells were cross-linked with 1% formaldehyde for 10 min at room temperature and quenched with glycine to a final concentration of 0.125 M for another 10 min. Chromatin was sonicated with a Bioruptor (Diagenode), cleared by centrifugation, and incubated overnight at 4 °C with 5–7 µg of the desired antibodies: anti-DDX21 (Novus Biologicals BP100-1781 and NBPI-83310), anti-Pol II Ser2p (Active Motif 61084), anti-Pol II (Santa Cruz Biotechnology sc-899), anti-CTCF (Cell Signaling 2899) anti-CDK9 (1:1 mix of Santa Cruz Biotechnology sc-8338 and sc-484). Immunocomplexes were immobilized with 100 µl of protein-G Dynal magnetic beads (Life Technologies) for 4 h at 4 °C, followed by stringent washes and elution. Eluates were reverse cross-linked overnight at 65 °C and deproteinized with proteinase K at 56 °C for 30 min. DNA was extracted with phenol chloroform, followed by ethanol precipitation. ChIP-seq libraries were prepared according to the NEBNext protocol and sequenced using Illumina HiSeq 2500. ChIP-qPCR analyses were performed in a Light Cycler 480II machine (Roche). ChIP-qPCR signals were calculated as percentage of input. Fold induction was calculated over a negative genomic region. All primers used in qPCR analyses are shown in Supplementary Table 1. All ChIP antibodies have been previously validated unless otherwise specified.

ChIP-seq analyses. Sequences were mapped using DNAnexus software tools and analysed by QuEST and MACS2. For QuEST, ChIP-seq peaks were determined using a kernel density estimate bandwidth of 30, a ChIP candidate threshold of 20, a ChIP extension fold enrichment of 3, and a ChIP-to-background fold enrichment of 3. Wig files were generated with QuEST and used for visualization in the UCSC Genome Browser and for obtaining average signal profiles. Average ChIP-seq signal profiles around the centre of DDX21 ChIP-seq peaks were generated with the Sitepro tool, which is part of the Cistrome/Galaxy pipeline. We used the HOMER software to associate DDX21 ChIP-seq peaks to different genomic features. Functional annotation and Gene Ontology categories were obtained with GREAT³⁴. For ascribing DDX21 binding to snoRNA-host genes, we generated a file containing all snoRNAs and their associated host genes and only snoRNAs residing within introns of RefSeq genes were used in this study. To avoid redundancies all entries were inspected manually in the UCSC genome browser.

All genomic data sets have been deposited under the GEO record GSE56802. Other data sets used in this study were obtained from GSM891237, GSE36620, GSM1249888, GSM1249889, GSE20598, GSM1249897 and GSE20598.

For mapping DDX21 ChIP-seq to the rDNA, we obtained the DNA consensus sequence of the 43-kb ribosomal locus NCBI (GeneBank ID: U13369.1). This 43 kb itself is unique relative to other locations in the human genome; however, as noted, it is repeated hundreds of times in each of the 5 chromosomal clusters. Using this

unique 43-kb region, we used the Bowtie algorithm to map ChIP-seq reads with standard parameters used for mapping to the Hg19 human genome build. The same strategy has been employed by other groups to map transcription factors ChIP reads to the rDNA locus⁷.

RNA extraction and qRT-PCR. RNA was isolated using Trizol (Life Technologies) according to the manufacturer's protocol. All RNA samples were DNase-treated with the Turbo DNA-Free kit (Ambion). cDNA was generated using SuperScript VILO (Life Technologies) according to manufacturer instructions. qPCR analyses were performed on the Light Cycler 480II (Roche). All primers used are shown in Supplementary Table 1.

Immunofluorescence. HEK293 cells were seeded into 24-well plates containing 12-mm glass coverslips and cultured for 16 h in DMEM containing 10% FBS (v/v). Cells were then treated for the indicated drugs (refer to the corresponding figure legends for drug concentration and time scale of the experiment). Cells were fixed in 4% paraformaldehyde for 10 min at room temperature, 3 × 5-min washes with PBS, followed by an ice-cold methanol fix for 2 min and 2 × 5-min washes with PBS. Cells were permeabilized in PBS containing 0.3% (v/v) Triton X-100 for 5 min, and blocked overnight at 4 °C in PBT buffer (PBS with 1% BSA, 0.1% Triton X-100 (v/v), 0.05% sodium azide (w/v)). After blocking, coverslips were incubated in PBT with the corresponding antibody. For DDX21 (Novus Biologicals NBPI-83310) the antibody was diluted 1:200 and incubated at room temperature for 2 h. For fibrillarin (Cell Signaling C13C3) the antibody dilution was 1:100 and incubated at room temperature for 2 h. Coverslips had 3 × 5-min washes with PBT and incubated with the Alexa-Fluor 568 secondary antibody (1:1,000; Life Technologies) for 1 h. Cells were washed 3 × 5 min with PBT, 2 × 5 min with PBS, rinsed briefly with water and mounted onto glass slides using VECTASHIELD mounting medium with DAPI. All images were taken and processed using a Zeiss LSM700 confocal microscope.

Western blots and co-immunoprecipitation. HEK293 nuclear extracts were prepared as described previously³⁵. For immunoprecipitations, extracts were incubated overnight with 3 µg of the desired antibody pre-bound to protein G-sepharose (Pierce). In some case protein extracts were treated with RNaseA (20 µg ml⁻¹). Immunocomplexes were eluted in 2× Laemmli buffer and resolved in an 8% acrylamide gel. For western blots the following antibodies were used according to manufacturer instructions: anti-NOP58 (Bethyl A302-718A); anti-fibrillarin (Cell Signaling C13C3); anti-DKC1 (Gene Tex GTX109000); anti-Flag (Sigma); anti-DDX21 (Novus Biologicals NB100-1781); anti-LARP7 (a gift from D. H. Price); anti-CDK9 (Santa Cruz Biotechnology sc-484); anti-cyclinT1 (Santa Cruz Biotechnology sc-10750); and anti-HEXIM1 (Bethyl A303-113A). All antibodies have been previously validated unless otherwise specified.

iCLIP and data analysis. The iCLIP method was performed as described before with the specific modifications below^{15,36}. Twenty-four hours after dox treatment (0.025 µg ml⁻¹), FH-DDX21^{WT} or FH-DDX21^{SAT} HEK293 cell lines were ultraviolet crosslinked to a total of 0.3 J cm⁻². Whole-cell lysates were generated in CLIP lysis buffer (50 mM HEPES, 200 mM NaCl, 1 mM EDTA, 10% glycerol, 0.1% NP-40, 0.2% Triton X-100, 0.5% N-lauroylsarcosine) and briefly sonicated using a probe-tip Branson sonicator to solubilize chromatin. Each iCLIP experiment was normalized for total protein amount, typically 2 mg, and partially digested with RNaseA (Affymetrix) for 10 min at 37 °C and quenched on ice. FH-DDX21 was isolated with anti-Flag agarose beads (Sigma) for 3 h at 4 °C on rotation. Samples were washed sequentially in 1 ml for 5 min each at 4 °C: 2× high stringency buffer (15 mM Tris-HCl, pH 7.5, 5 mM EDTA, 2.5 mM EGTA, 1% Triton X-100, 1% sodium deoxycholate, 120 mM NaCl, 25 mM KCl), 1× high salt buffer (15 mM Tris-HCl pH 7.5, 5 mM EDTA, 2.5 mM EGTA, 1% Triton X-100, 1% sodium deoxycholate, 1 M NaCl), 1× NT2 buffer (50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 1 mM MgCl₂, 0.05% NP-40). Purified FH-DDX21 was then eluted off anti-Flag agarose beads using competitive Flag peptide elution. Each sample was resuspended in 500 µl of Flag elution buffer (50 mM Tris-HCl, pH 7.5, 250 mM NaCl, 0.5% NP-40, 0.1% sodium deoxycholate, 0.5 mg ml⁻¹ Flag peptide) and rotated at 4 °C for 30 min. The Flag elution was repeated once for a total of 1 ml elution. FH-DDX21 was then captured using anti-haemagglutinin agarose beads (Pierce) for 1 h at 4 °C on rotation. Samples were then washed as previously in the anti-Flag agarose beads. 3'-end RNA dephosphorylation, 3'-end single-stranded RNA ligation, 5' labelling, SDS-PAGE separation and transfer, autoradiograph, ribonucleoprotein isolation, proteinase K treatment, and overnight RNA precipitation took place as previously described¹⁵. The 3'-end single-stranded RNA ligation adaptor was modified to contain a 3' biotin moiety as a blocking agent (Supplementary Table 1). The iCLIP library preparation was performed as described elsewhere (R.A.F., L.M., R.C.S. and H.Y.C., unpublished observations). Final library material was quantified on the BioAnalyzer High Sensitivity DNA chip (Agilent) and then sent for deep sequencing on the Illumina NextSeq machine for 1 × 75-bp cycle run. iCLIP data analysis was performed as previously described¹⁵. For analysis of repetitive non-coding RNAs, custom annotation files were built from the Rfam database and reads were mapped under standard iCLIP processing steps. For the repetitive RNA analysis we normalized

the iCLIP reverse transcription stops across each RNA transcript, which allows the comparison of the shape of the profile, with this normalization, the y axis is informative only for the relative binding preference and does not imply the strength of binding.

siRNA and antisense oligonucleotide knockdown. For *DDX21* knockdown, HEK293 cells (2.5×10^5) were transfected in DMEM supplemented with 5% FBS without antibiotics using RNAiMAX (Life Technologies). *DDX21* (20 nM) or control siRNA was used in this study. Notably, *DDX21* depletion was difficult to achieve, thus for efficient *DDX21* knockdown (60–80% at the protein level) three consecutive siRNA transfections were required. siRNA diced pools were generated in J.W.'s laboratory using recombinant *Giardia lamblia* Dicer.

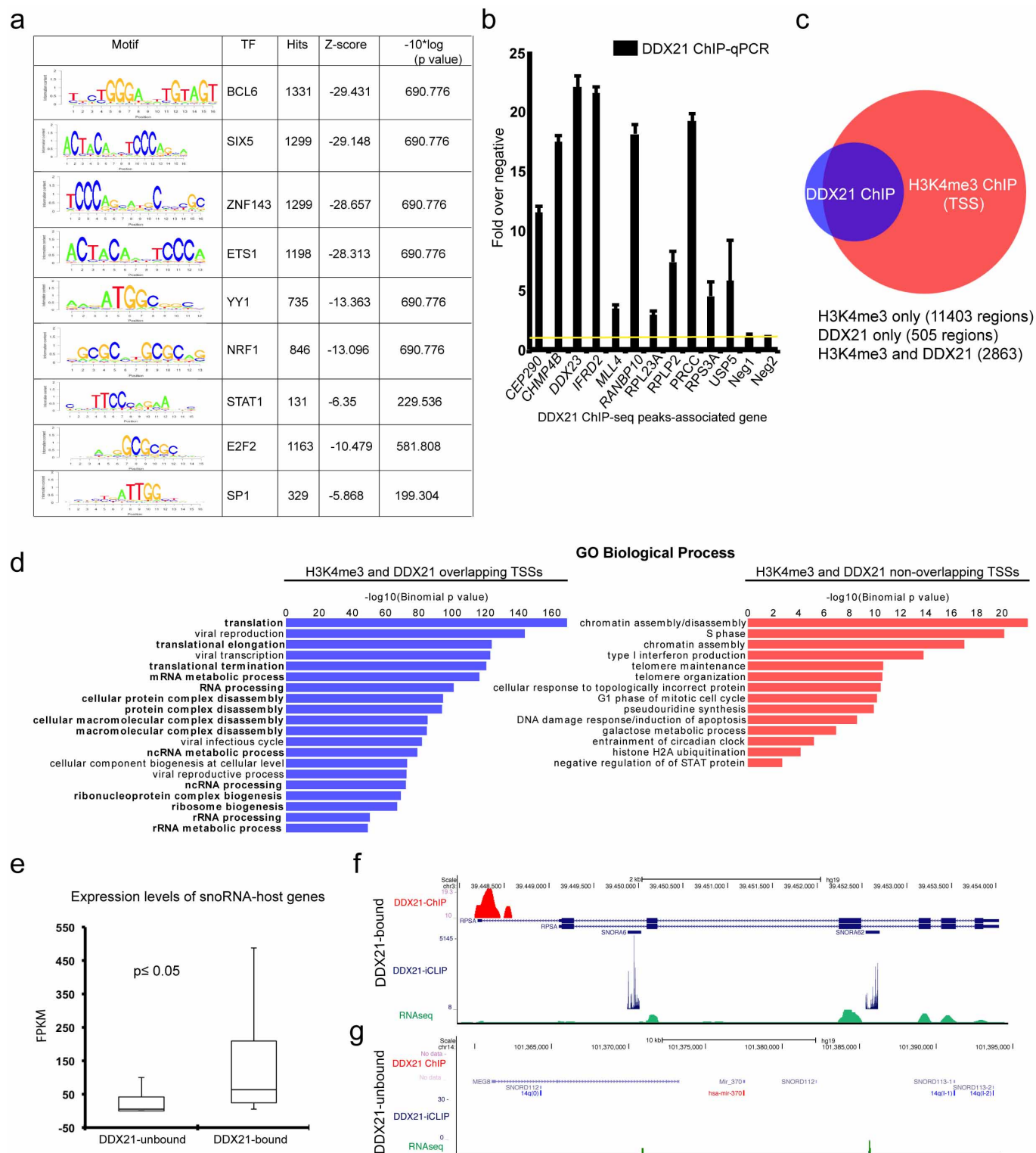
ASO depletion of 7SK was performed as previously described²⁶. In brief, HEK293 cells (2×10^6) were nucleofected with the Amaxa Nucleofector 2b (Lonza) with 1 nmol of scramble control, 5'-7SK, or 3'-7SK (ref. 26). Cells were cultured for 12 h at 37 °C and collected for ChIP-qPCR as described above. A fraction of each sample was collected with Trizol for assaying 7SK knockdown by qRT-PCR as described above.

Site-directed RNaseH rRNA cleavage assay. HEK293 cells were seeded and transfected with *DDX21* or control siRNA as described above. Twelve hours after the second siRNA transfection, *DDX21*^{WT} or *DDX21*^{SA7} cDNAs were induced by adding $0.25 \mu\text{g ml}^{-1}$ dox, after which cells were collected by scrapping the monolayer in ice-cold PBS. Total cellular RNA was isolated by Trizol extraction and RNeasy column clean up. To evaluate the fraction of methylation of specific nucleotides within the rRNA quantitatively, we optimized an established assay using RNaseH to cleave unmethylated RNA selectively²². Several sites were selected based on iCLIP reverse transcription stops to both the rRNA region of interest as well as reverse transcription stops on the targeting snoRNA transcript. Chimaeric 2'-Ome/DNA oligonucleotides were ordered for three sites (Supplementary Table 1) each containing three DNA nucleobases for RNaseH targeting. For each cleavage reaction 1 μg of total RNA was mixed with 0.5 pmol of a specific chimaeric oligonucleotide (final volume of 6 μl) and annealed to the rRNA by incubating at 80 °C for 2 min and then step-cooling the sample to 25 °C, decreasing the temperature 1 °C per second. Then 2.5 μl of 4 \times RNaseH cleavage buffer (80 mM Tris-HCl, pH 7.5, 40 mM MgCl₂, 400 mM KCl, 0.4 mM dithiothreitol, 20 mM sucrose), 1 μl of RNaseH (Roche, note: RNaseH from this supplier is critical as other suppliers or isolates have different specificity for site-directed cleavage), and 1 μl of SUPERaseIn (Life Technologies) were added to each reaction and incubated for 25 min at 37 °C. Samples were subsequently purified using RNeasy columns and eluted in 100 μl of water. For visualization and

quantification, 5 μl of each sample was mixed with 5 μl of GLBII (Life Technologies) and heated to 65 °C before agarose gel electrophoresis. Full-length and cleavage products were imaged and quantified on a ChemiDoc XRS+ (BioRad).

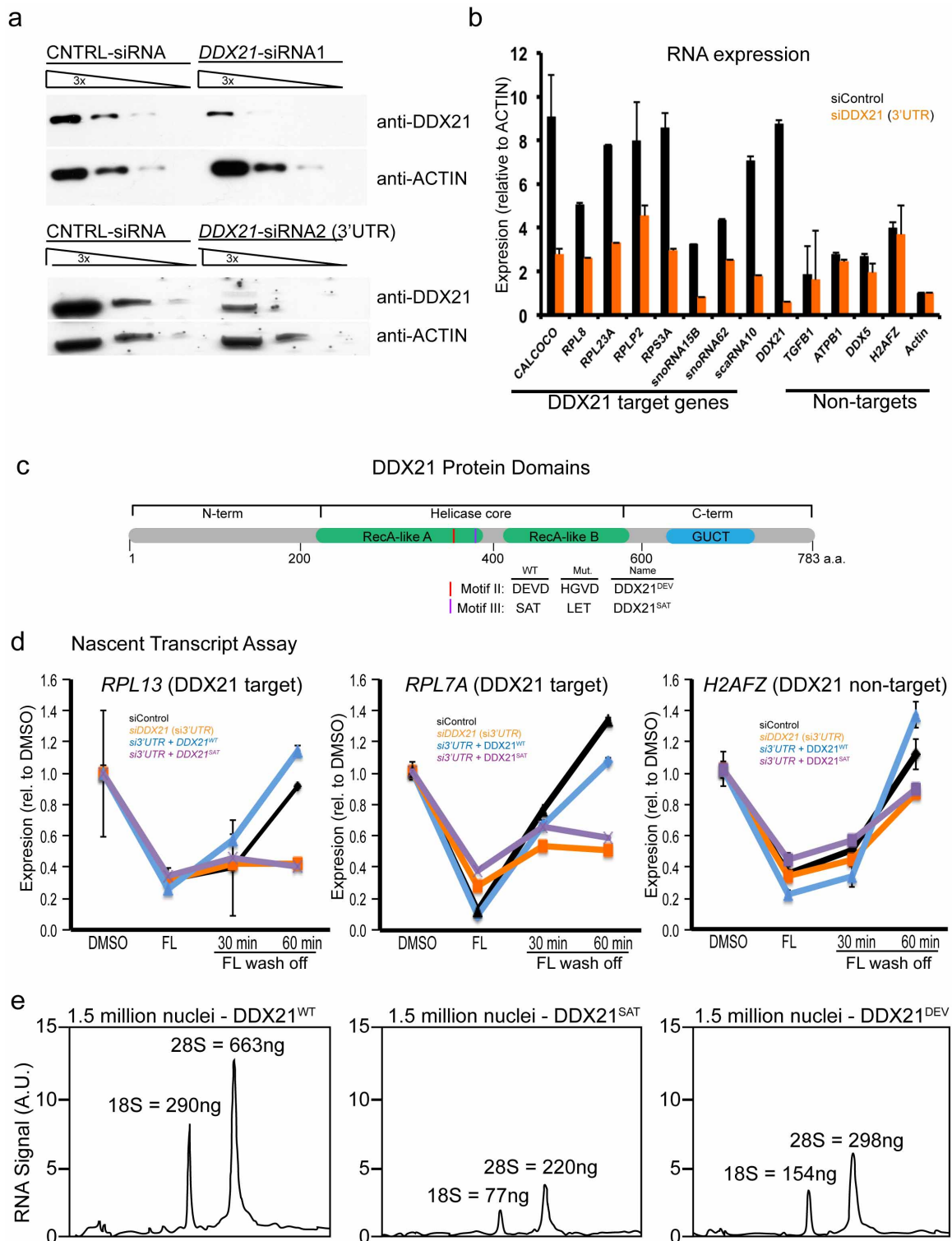
P-TEFb release assay. The P-TEFb release assay was performed as previously described³⁷ with some modifications. Five micrograms of anti-HEXIM1 (ab25388) antibody was pre-bound to protein A Dynal magnetic beads (Life Technologies) and incubated with 2.5 mg of HeLa cell nuclear extracts to immobilize the inactive 7SK snRNP complex. The resulting immunocomplexes were incubated with increasing amounts of purified Flag-DDX21 and incubated for 2 h on ice. A magnetic separator was used to sequester the remaining HEXIM1-bound 7SK snRNP, and the resulting eluates were collected and analysed by western blotting. For purification of Flag-DDX21, nuclear extracts from HEK293 cells stably expressing Flag-DDX21 were prepared using a modified version of ref. 35. Purified nuclei were extracted in Dignam and Roeder buffer C and cleared by centrifugation. The salt concentration of the cleared extracts was adjusted to 250 mM with Dignam and Roeder buffer D. The resulting nuclear extracts were incubated with Flag-M2 agarose beads (Sigma) for 2 h to immobilize Flag-DDX21, followed by stringent washes: 5 \times 2 min with Dignam and Roeder buffer C at an NaCl concentration of 500 mM and 1 \times 5 min with buffer D-C (20 mM HEPES, 20% glycerol, 0.1 mM EDTA, 150 mM NaCl, 0.75 mM MgCl₂ and 0.05 M KCl). Flag-DDX21 was eluted with the 3 \times Flag peptide (Sigma) in buffer D-C.

31. Flynn, R. A., Almada, A. E., Zamudio, J. R. & Sharp, P. A. Antisense RNA polymerase II divergent transcripts are P-TEFb dependent and substrates for the RNA exosome. *Proc. Natl Acad. Sci. USA* **108**, 10460–10465 (2011).
32. Wassarman, D. A. & Steitz, J. A. Structural analyses of the 7SK ribonucleoprotein (RNP), the most abundant human small RNP of unknown function. *Mol. Cell. Biol.* **11**, 3432–3445 (1991).
33. Sharma, A. *et al.* The Werner syndrome helicase is a cofactor for HIV-1 long terminal repeat transactivation and retroviral replication. *J. Biol. Chem.* **282**, 12048–12057 (2007).
34. McLean, C. Y. *et al.* GREAT improves functional interpretation of cis-regulatory regions. *Nature Biotechnol.* **28**, 495–501 (2010).
35. Carey, M. F., Peterson, C. L. & Smale, S. T. Dignam and Roeder nuclear extract preparation. *Cold Spring Harb. Protoc.* **2009**, 10.1101/pdb.prot5330 (2009).
36. König, J. *et al.* iCLIP-transcriptome-wide mapping of protein-RNA interactions with individual nucleotide resolution. *J. Vis. Exp.* **50**, 2638 (2011).
37. Krueger, B. J., Varzavand, K., Cooper, J. J. & Price, D. H. The mechanism of release of P-TEFb and HEXIM1 from the 7SK snRNP by viral and cellular activators includes a conformational change in 7SK. *PLoS ONE* **5**, e12335 (2010).



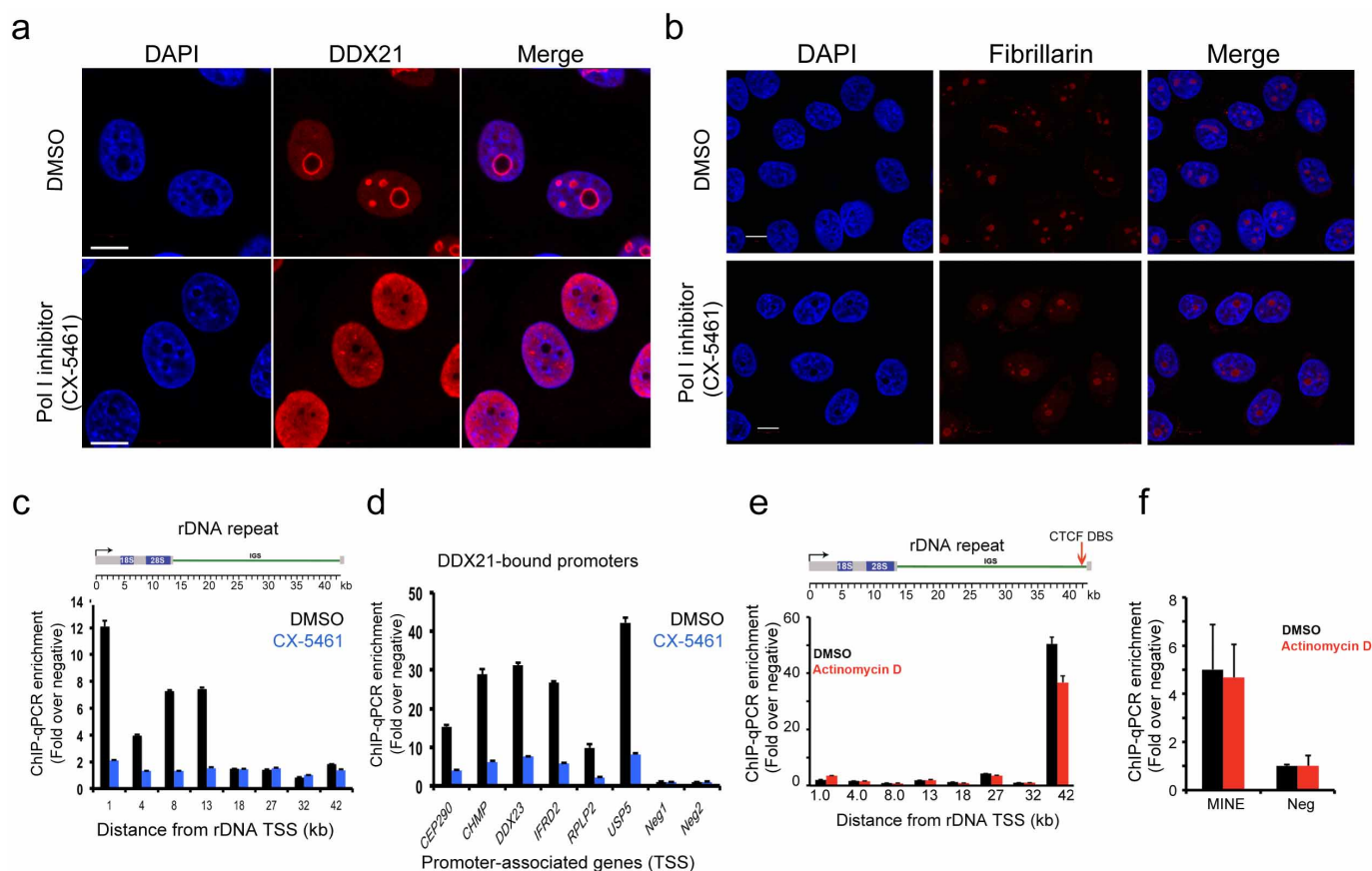
Extended Data Figure 1 | DDX21 associates with non- and protein-coding ribosomal genes. **a**, MEME analysis of DDX21-bound regions defined by DDX21 ChIP-seq. Motif logo, annotated transcription factor, number of motif instances within the ChIP-seq regions, Z score, and P value for each motif are shown. **b**, DDX21 ChIP-qPCR from HeLa cell chromatin extracts with primers spanning a representative number of loci found to be enriched in the DDX21 ChIP-seq analyses from HEK293 cells. Data are mean and s.d. of three independent experiments. **c**, Comparison of DDX21 (this study) and H3K4me3 (publicly available data, see Methods for accession numbers) ChIP-seq-bound regions. 2,863 regions are common between the data sets, 505 regions are unique to DDX21, and 11,403 regions are unique to the

H3K4me3 data set. **d**, Gene Ontology terms for H3K4me3 regions that are either DDX21-bound (left) or not bound by DDX21 (right). **e**, Box plots representing the expression levels of snoRNA-host genes whose promoter regions are either bound or not by DDX21. As shown, snoRNA-host gene promoters bound by DDX21 are, on average, more highly expressed than those not occupied by DDX21. Fragments per kilobase of exon per million mapped reads (FPKM) values were taken from publicly available HEK293 RNA-seq data (see Methods for accession number). The P value ($P \leq 0.05$) was calculated using the Wilcoxon signed-rank test. **f**, **g**, UCSC genome browser tracks depicting DDX21 ChIP-seq and iCLIP-seq, and RNA-seq enrichment profiles at differentially expressed snoRNA-host genes in HEK293 cells.



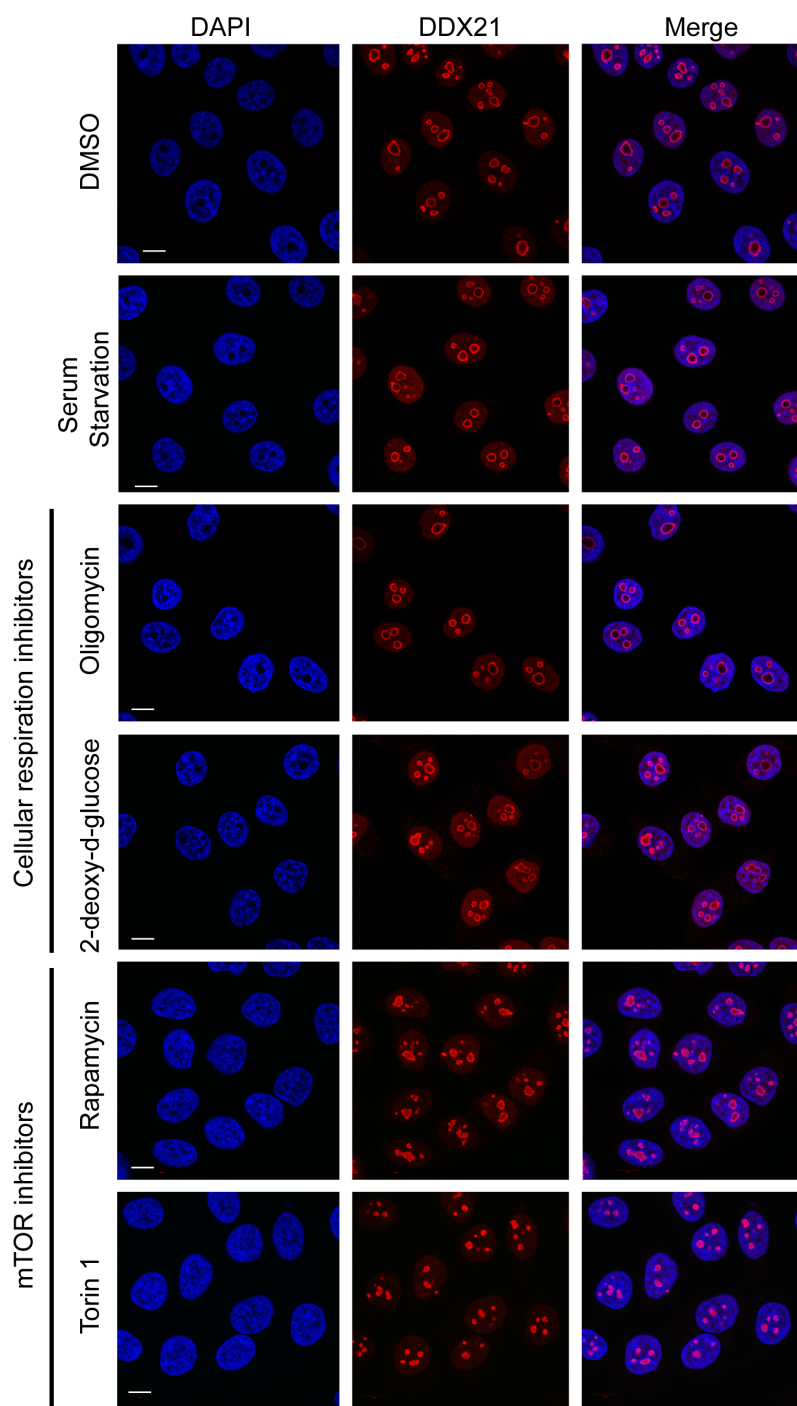
Extended Data Figure 2 | DDX21 positively regulates transcription of Pol I- and Pol II-dependent ribosomal genes. **a**, siRNA-mediated knockdown of the DDX21 antibody used for ChIP. We transfected HEK293 cells with two different sets of siRNAs targeting endogenous *DDX21* mRNA (siRNA1 and siRNA2 (3' UTR)) and performed western blots with the indicated antibodies. As shown, the DDX21-specific band is diminished in cells transfected with *DDX21*-targeting siRNAs, but not with control siRNAs. Actin was used as a loading control for this experiment. **b**, RT-qPCR analysis assessing the RNA expression levels of the same genes analysed in Fig. 1h upon *DDX21* knockdown by a second siRNA that targets the 3' UTR of *DDX21* mRNA. Data are mean and s.d. of three independent experiments. For *DDX21*-target genes the difference between control and *DDX21* siRNA is significant,

$P \leq 0.05$ (Student's *t*-test). **c**, Diagram of DDX21 protein domains. The two conserved RecA-like (A and B) domains and the GUCT domains are shown in green and blue, respectively. Amino acids targeted for mutation¹² to convert *DDX21*^{WT} into *DDX21*^{SAT}, the ATP-hydrolysis mutant, are indicated with red and purple lines in the diagram. Specific amino acid changes are displayed below. **d**, qRT-PCR analysis assessing nascent unspliced mRNA levels from additional *DDX21*-target and *DDX21*-non-target promoters. For a detailed description see Fig. 1j. Data are mean and s.d. of three biological replicates. **e**, Nuclear rRNA abundance analysis by RNA BioAnalyzer of HEK293 cells depleted of DDX21 and rescued with *DDX21*^{WT}, *DDX21*^{SAT}, or *DDX21*^{DEV}. For each analysis, total RNA was isolated from 1,500,000 nuclei. Total nanogram amounts are shown for each of the two large rRNA subunits.



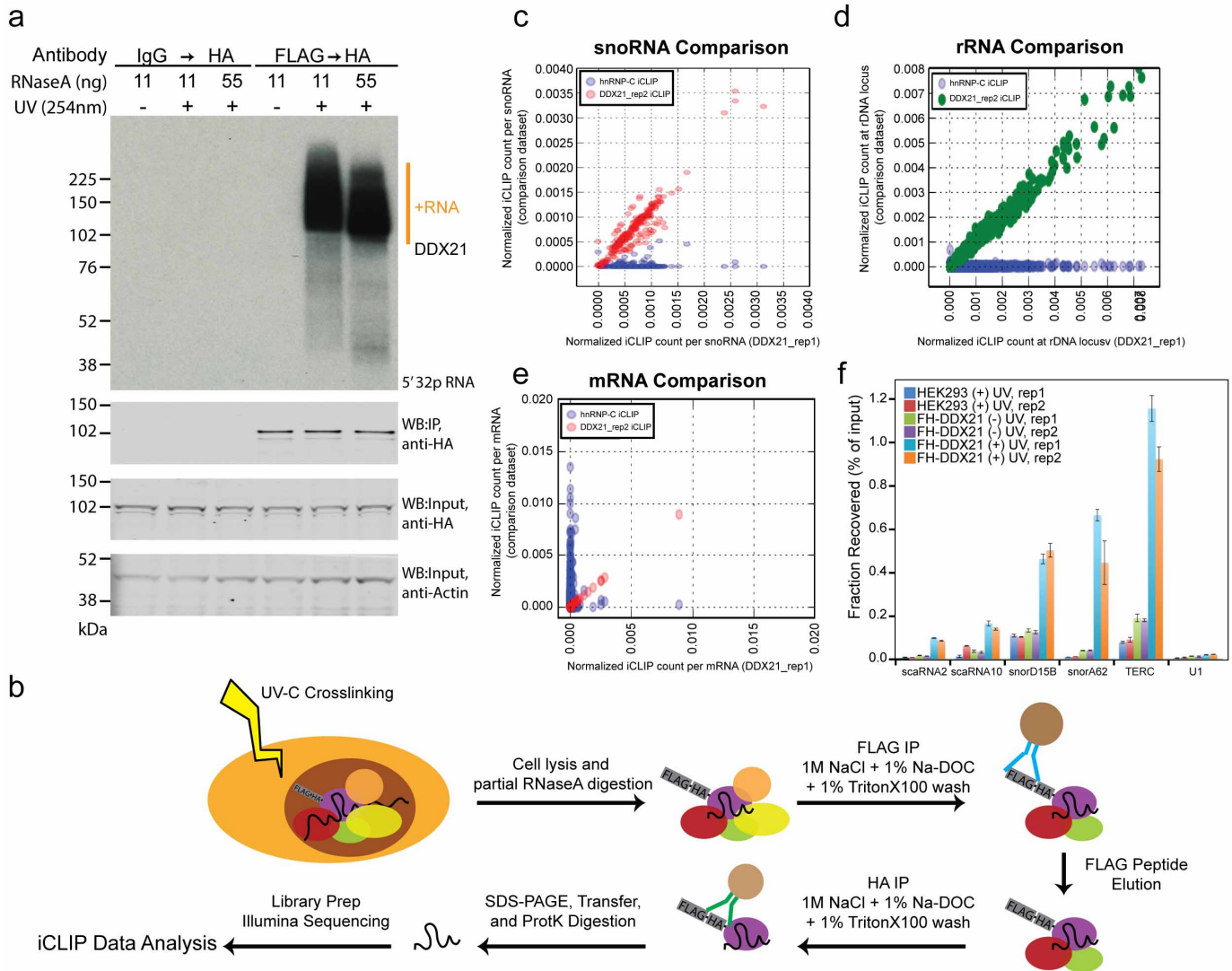
Extended Data Figure 3 | Selective inhibition of Pol I alters DDX21 nuclear localization and chromatin association. **a, b**, Immunofluorescence images of methanol-fixed HEK293 cells after 1 h incubation with either DMSO or 2 μ M of the specific Pol I inhibitor CX-5461. DDX21 (**a**) and fibrillarin (**b**) immunolabellings are shown. Scale bars, 10 μ m. **c, d**, ChIP-qPCR analyses from HEK293 sampling DDX21 genomic occupancy, at the rDNA locus (**c**) and at a representative panel of Pol II-regulated gene promoters (**d**), after treatment with DMSO or CX-5461. Data are mean and s.d. of three independent experiments. As displayed, inhibition of Pol I alters DDX21 nuclear localization

and this coincides with nearly complete eviction of DDX21 from Pol I- and Pol II-regulated genes. **e, f**, ChIP-qPCR analyses from HEK293 cells treated with 50 ng ml⁻¹ of actinomycin-D for 1 h. Binding of the transcriptional repressor CTCF across the rDNA locus (**e**) and the c-MYC insulator element (MINE) (**f**) demonstrates that actinomycin-D treatment does not effect CTCF binding to chromatin. Red arrow indicates relative location of the CTCF DNA-binding site (DBS) at the rDNA locus. Data are mean and s.d. of three independent experiments.



Extended Data Figure 4 | DDX21 nuclear re-localization is preferentially sensitive to acute transcriptional inhibition over other cellular stressors. Immunofluorescence analyses of HEK293 cells after targeting different metabolic pathways. For inhibition of mitogen, cells were starved for 16 h in the

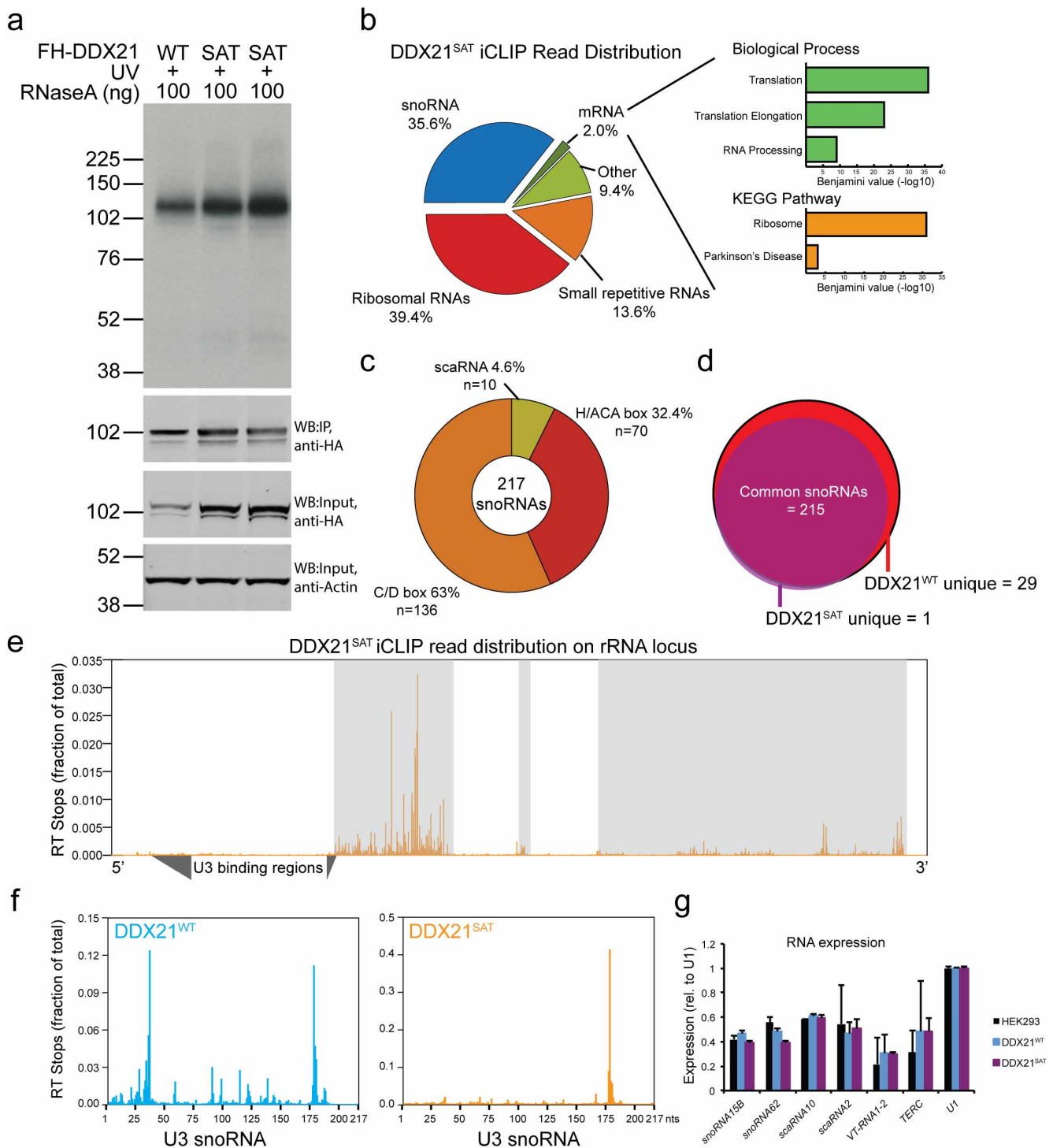
absence of serum. For cellular respiration inhibition, cells were treated for with either oligomycin (100 μ M) or 2-deoxy-D-glucose (10 mM) for 1 h. To inhibit the mTOR pathway, cells were treated with 250 nM of either Torin 1 or rapamycin for 2 h.



Extended Data Figure 5 | Tandem affinity iCLIP of FH-DDX21^{WT}.

a, FH-DDX21^{WT} iCLIP ³²P-autoradiogram and western blots. All samples were loaded with constant input lysate amounts (actin loading). FH-DDX21^{WT} was isolated from HEK293 cells induced to express the transgene for 24 h and crosslinked with ultraviolet light (top panel same as Fig. 3a). **b**, Schematic of the modified iCLIP procedure. To achieve high stringency and specificity Flag-HA-DDX21^{WT} is first purified on anti-Flag-M2 agarose beads, washed with 1 M NaCl, 1% Triton X-100 and 1% sodium deoxycholate. Complexes are specifically eluted with Flag peptide and recaptured with anti-HA agarose. Standard iCLIP steps were performed thereafter to generate deep sequencing libraries. **c–e**, Scatter plot analysis of iCLIP reverse transcription stops on

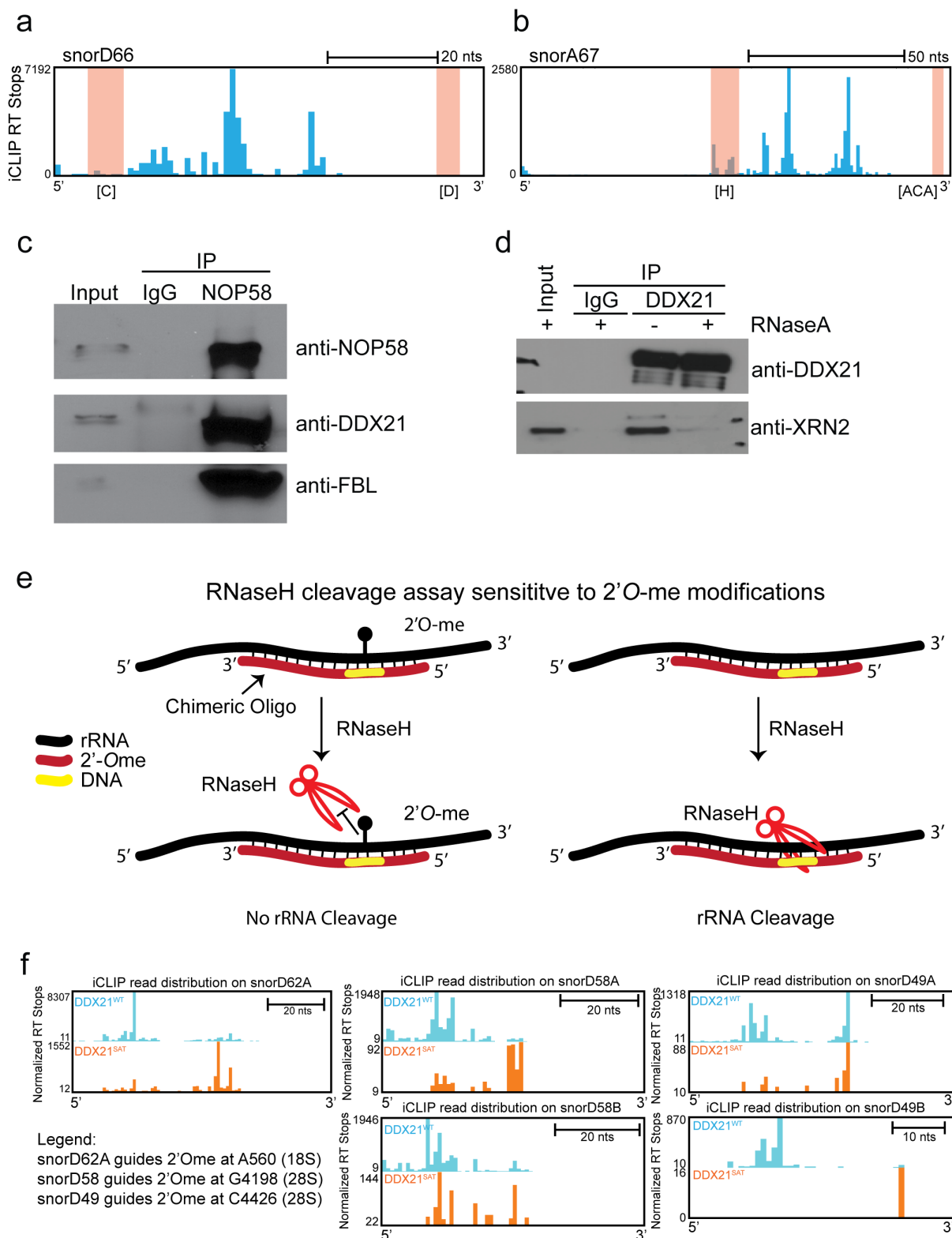
snoRNAs, rRNA and mRNAs within the FH-DDX21^{WT} (this study) and hnRNP-C (ref. 16; publicly available data) data sets. Little concordance between the data sets is evident, suggesting specific transcriptome targets of these two RNA binding proteins (RBPs). **f**, DDX21^{WT} ultraviolet RNA immunoprecipitation qRT-PCR of FH-DDX21^{WT} was performed in three conditions: native HEK293 cells crosslinked with ultraviolet light; FH-DDX21^{WT} HEK293 cells without crosslinking; and FH-DDX21^{WT} HEK293 cells with ultraviolet crosslinking. snoRNAs, scaRNAs and TERC were validated targets identified in the sequencing data. Each experiment was performed in biological duplicates (rep1 and rep2) and error bars represent s.d. of technical triplicates.



Extended Data Figure 6 | Tandem affinity iCLIP of FH-DDX21^{SAT}.

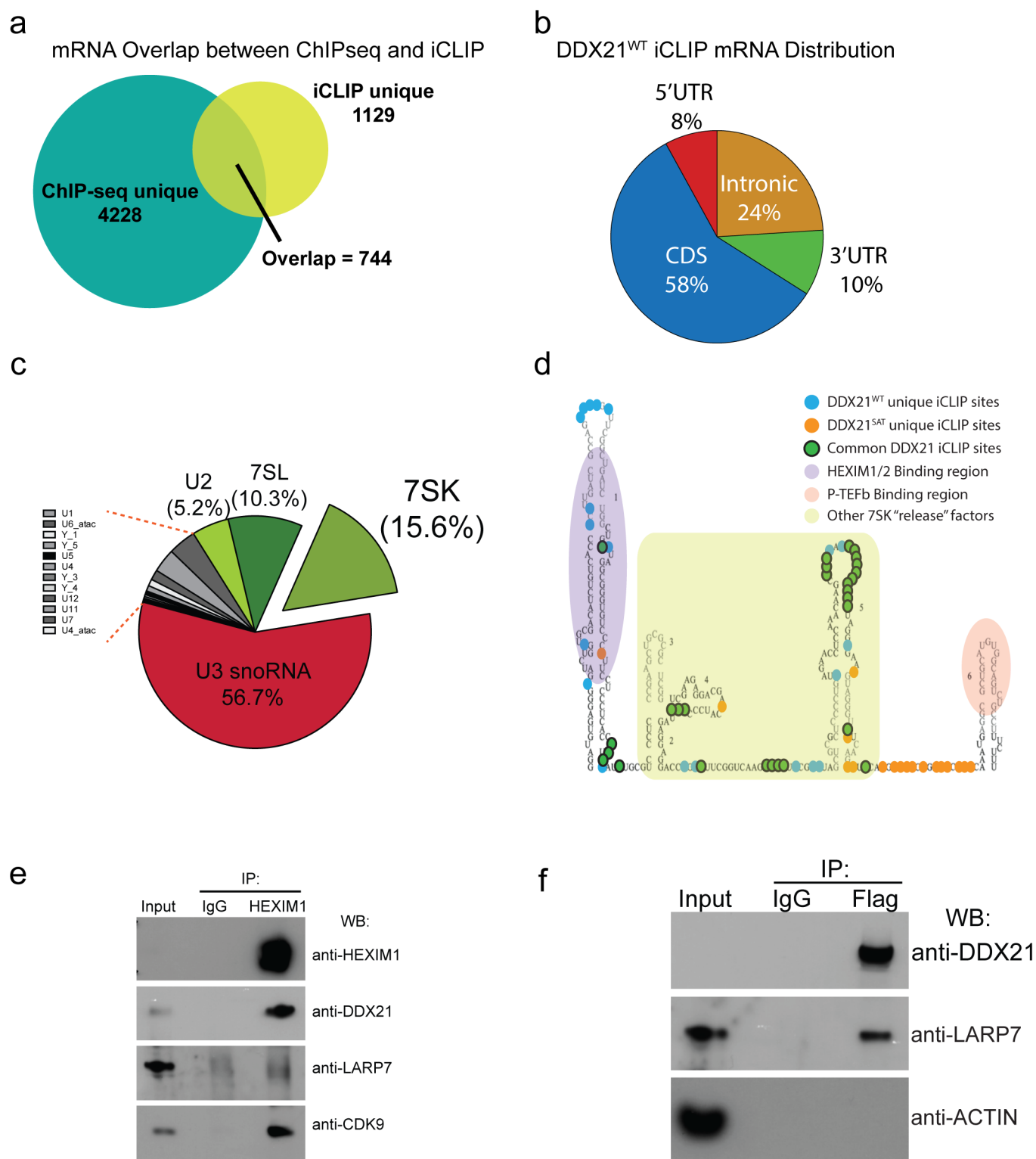
a, FH-DDX21^{SAT} was isolated from HEK293 cells induced to express the transgene for 24 h, at which point we did not observe significant dominant negative effects. iCLIP was performed as described for FH-DDX21^{WT}, and biological duplicates of FH-DDX21^{SAT} iCLIP ³²P-autoradiogram and western blots (lanes 2 and 3) are shown. All samples were loaded with constant input lysate amounts (actin loading). FH-DDX21^{WT} was loaded as a control. WB, western blot. **b**, Left, DDX21^{SAT} iCLIP reads annotated to known repetitive (rRNA and snRNAs) and non-repetitive (hg19 genome build: mRNAs and snRNAs) regions of the human genome. Categories are notes with their respective percentage of the total iCLIP experiment. Right, enriched Gene Ontology and KEGG pathway terms from DDX21^{SAT}-bound mRNAs obtained using the DAVID tool. The *x* axis values (in log scale) correspond to the negative Benjamini *P* value. **c**, Distribution of all DDX21^{SAT}-bound snoRNAs,

representing C/D box, H/ACA box and scaRNAs. The number (*n*) and fraction (per cent) of each snoRNA type is displayed. **d**, Comparison of the snoRNAs bound by DDX21^{WT} and DDX21^{SAT}, revealing significant overlap between the active and catalytically inactive DDX21. **e**, DDX21^{SAT} iCLIP reads mapped to the transcribed region of the rDNA. **f**, DDX21^{WT} (left) and DDX21^{SAT} (right) iCLIP reads mapped to the repetitive U3 snoRNA. Binding is represented as reverse transcription stops per nucleotide normalized to the total number of reverse transcription stops mapping to the U3 snoRNA. Two strong binding sites are evident between bases 25–40 and 175–185 of U3 in DDX21^{WT} iCLIP, whereas the 5' binding site is reduced in DDX21^{SAT}. **g**, qRT-PCR analysis assessing the expression levels of several snoRNAs 24 h after expression of either DDX21^{WT} or DDX21^{SAT}. This experiment was performed in biological duplicates. Data are mean and s.d. of technical triplicates.



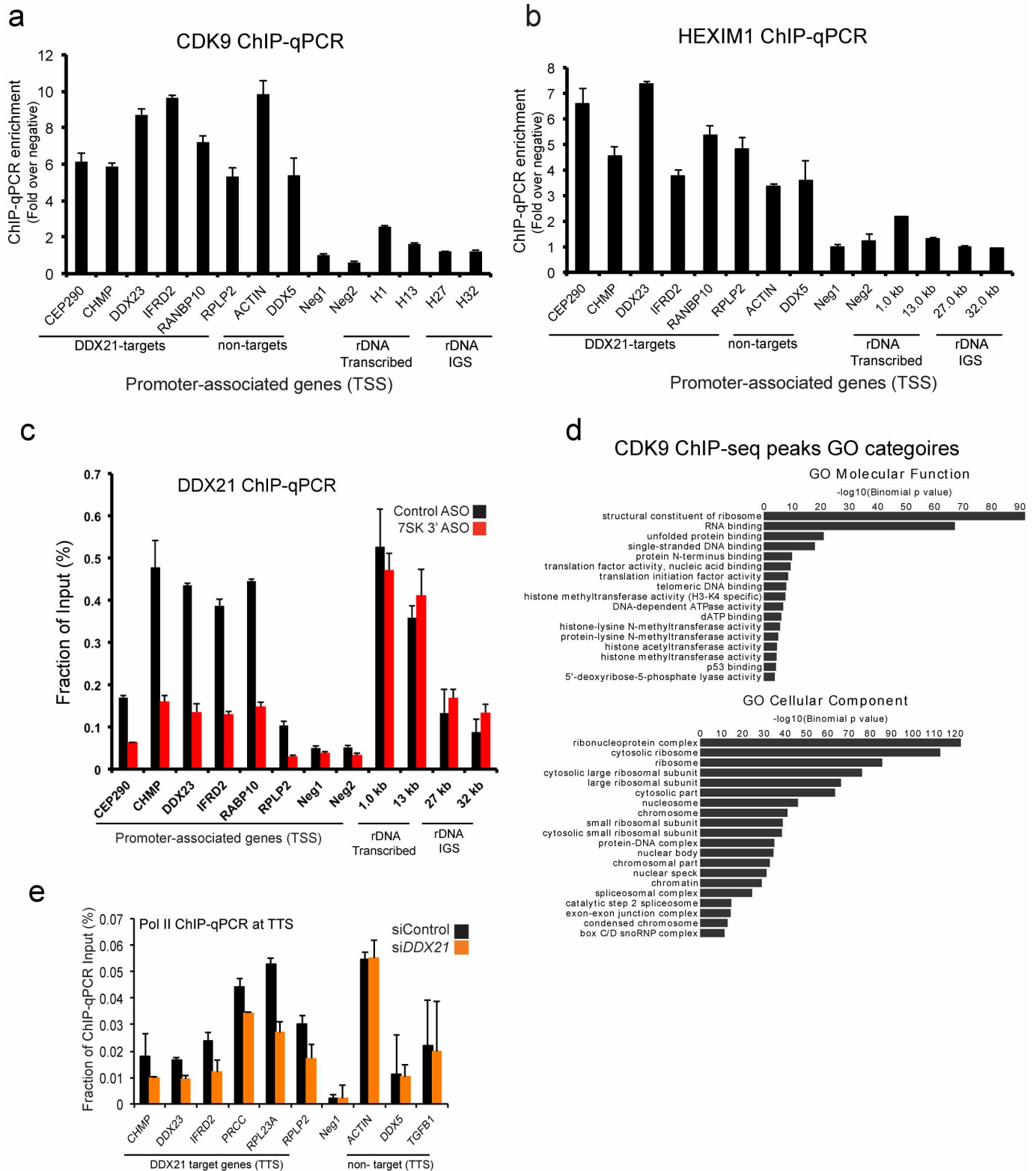
Extended Data Figure 7 | DDX21 functionally interacts with snoRNAs and the snoRNP. **a**, UCSC genome browser view of DDX21^{WT} iCLIP reads across the snorD66 snoRNA. The C box [C] and D box [D] regions are highlighted in red. **b**, Same visualization as in **a** but showing the snorA67 snoRNA with the H box [H] and ACA box [ACA] regions highlighted. **c**, Immunoprecipitation of NOP58 from HEK293 nuclear extracts confirms DDX21 as a protein member of the snoRNP machinery. As a control for this experiment we performed western blots against FBL, a well-known NOP58-interacting partner and an essential factor of the snoRNP machinery. **d**, DDX21 interacts with XRN2, a 5'–3' exoribonuclease required for maturation and processing of

snoRNAs. The DDX21–XRN2 interaction appears to be bridged by RNA, as treatment of the nuclear lysates with RNaseA abolishes the interaction. **e**, Schematic of the site-directed RNaseH cleavage of RNA sensitive to 2'-O-me. RNA of interest is hybridized to a 2'-O-me/DNA chimaeric oligonucleotide in which the DNA nucleotides specifically target the ability of RNaseH to interrogate the 2'-O-me status of a single nucleotide. 2'-O-me will inhibit RNaseH and leave intact RNA, while unmethylated RNA will be cleaved. **f**, UCSC genome browser view of DDX21^{WT} and DDX21^{SAT} iCLIP reads across the snoRNAs responsible for guiding the modifications tested in Fig. 3f.



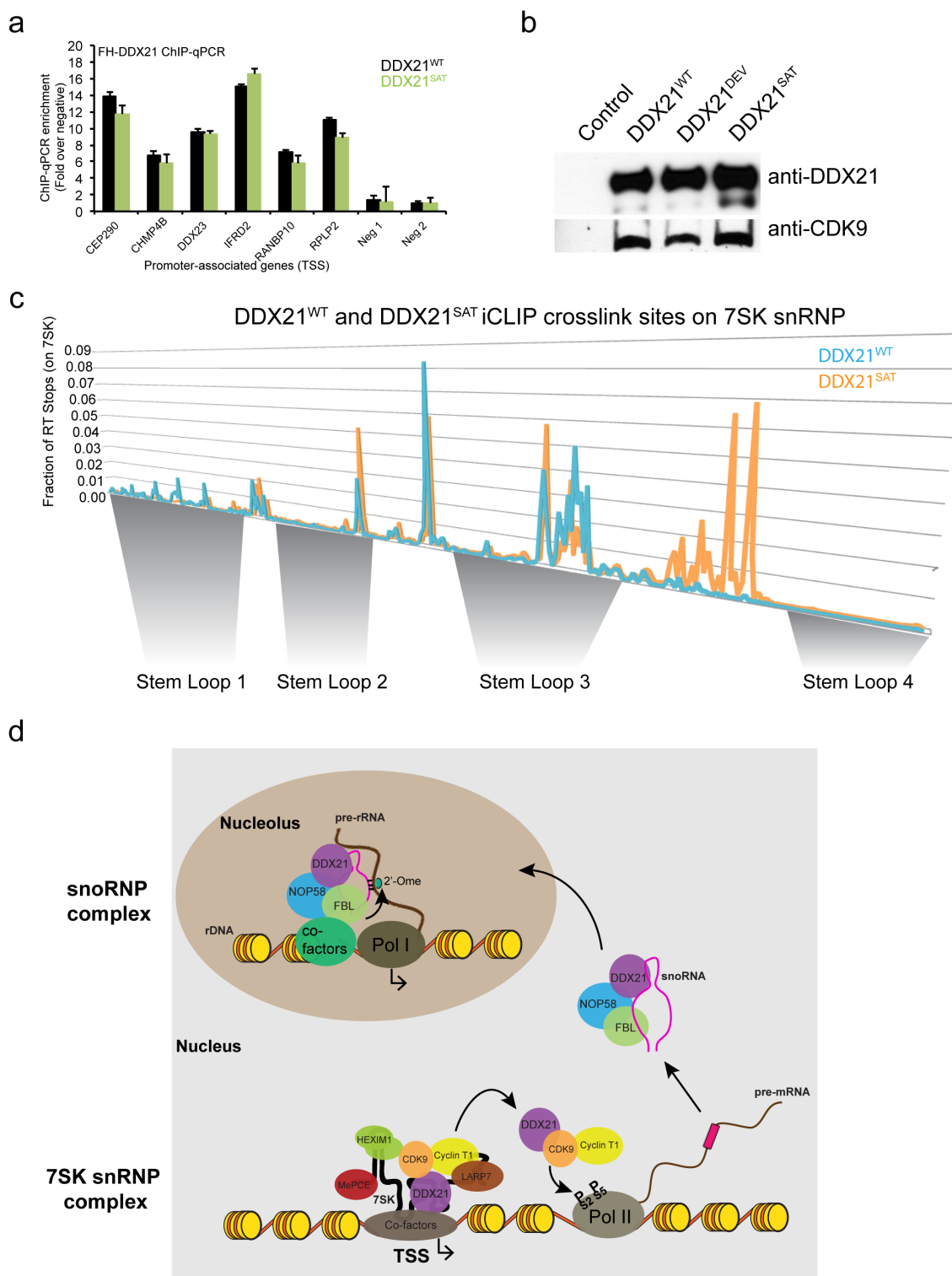
Extended Data Figure 8 | Association of DDX21 with the RNA and protein components of the 7SK snRNP. **a**, Comparison of DDX21^{WT} ChIP-seq targets to DDX21^{WT} iCLIP-bound mRNAs. The numbers of unique and common genes are represented, revealing that most ChIP-seq-bound genes are not immunoprecipitated by iCLIP but that some are recovered in both assays. **b**, iCLIP read distribution of DDX21^{WT}-target mRNAs categorized by the regions within mRNAs that were bound. Most iCLIP reads fell outside the 5' UTR. **c**, DDX21^{WT} iCLIP reads mapping to short repetitive RNAs of the human genome. Percentages of the top four short repetitive RNAs are shown. **d**, Secondary structure model of the 7SK snRNA annotated with iCLIP reverse

transcription stops identified from the DDX21^{WT} (blue) and DDX21^{SAT} (orange) experiments. Nucleotides commonly crosslinked are labelled in green. Known RNA binding protein sites: HEXIM1/2 is highlighted in purple; P-TEFb is highlighted in red; and other P-TEFb 'release' factors in the centre are highlighted in green. **e**, Co-immunoprecipitation analysis of HEXIM1 as assayed by western blotting for DDX21^{WT} and 7SK snRNP components (CDK9 and LARP7). **f**, Immunoprecipitation of Flag-HA-DDX21^{WT} from HEK293 nuclear extracts confirms DDX21 as a protein component of the 7SK snRNP through co-recovery of LARP7. The abundant protein actin, which is not part of the 7SK snRNP, was not recovered.



Extended Data Figure 9 | Binding of the DDX21-7SK snRNP at ribosomal gene promoters. **a**, **b**, ChIP-qPCR of CDK9 (**a**) and HEXIM1 (**b**) in HEK293 cells at representative Pol II-regulated, DDX21-target and -non-target gene promoters, negative control regions, and the rDNA locus. **c**, ChIP-qPCR of DDX21 in control or 3'-7SK-ASO-treated HEK293 cells at representative Pol II-regulated, DDX21-target gene promoters, negative control regions, and the rDNA locus. For the promoter-associated genes, $P \leq 0.05$ (Student's *t*-test)

when compared to control ASO. **d**, Gene Ontology molecular function and cellular component analysis of publicly available CDK9 ChIP-seq data³⁰. **e**, ChIP-qPCR of total Pol II in control or DDX21-targeting siRNA-treated HEK293 cells at representative TSSs of Pol II-regulated, DDX21-target gene promoters and negative control regions. Data are mean and s.d. of three independent experiments.



Extended Data Figure 10 | Catalytically inactive DDX21 is still incorporated into the 7SK snRNP. **a**, ChIP-qPCR of DDX21^{WT} (black) and DDX21^{SAT} (green) in HEK293 cells at representative TSSs of Pol II-regulated, DDX21-target gene promoters and negative control regions. Data are mean and s.d. of three independent experiments. **b**, Immunoprecipitation of DDX21^{WT}, DDX21^{DEV} or DDX21^{SAT} from HEK293 nuclear extracts confirms DDX21 interacts with CDK9 (P-TEFb) regardless of its catalytic activity. **c**, DDX21^{WT} (blue) and DDX21^{SAT} (orange) annotated iCLIP reads mapped across the 7SK snRNA. The four annotated stem-loops are marked below the graph. **d**, Model of multi-level control of ribosomal pathway by DDX21. In the nucleolus, DDX21 associates with the chromatin across the transcribed region of the rDNA and is a component of the snoRNP. Furthermore, DDX21

functionally interacts with the rRNA, snoRNAs and snoRNP to control 2'-Ome deposition on the rRNA in a helicase activity-dependent manner. In the nucleoplasm, DDX21 is bound to the promoter regions of ribosomal Pol II-transcribed genes, many of which contain precursor snoRNA transcripts. Mechanistically, DDX21 activates transcription of its target genes through the 7SK-P-TEFb axis. As part of the 7SK snRNP, DDX21 can facilitate the release of P-TEFb from the inhibitory complex in a manner dependent on ATP hydrolysis, leading to productive Pol II elongation and increased phosphorylation of Ser 2. Efficient transcription of its target genes enforces high expression of both snoRNAs and other ribosomal proteins critical for the rRNA maturation process, placing DDX21 as a central operator of the ribosomal pathway.

Mammalian polymerase θ promotes alternative NHEJ and suppresses recombination

Pedro A. Mateos-Gomez¹, Fade Gong², Nidhi Nair³, Kyle M. Miller², Eros Lazzerini-Denchi³ & Agnel Sfeir¹

The alternative non-homologous end-joining (NHEJ) machinery facilitates several genomic rearrangements, some of which can lead to cellular transformation. This error-prone repair pathway is triggered upon telomere de-protection to promote the formation of deleterious chromosome end-to-end fusions^{1–3}. Using next-generation sequencing technology, here we show that repair by alternative NHEJ yields non-TTAGGG nucleotide insertions at fusion breakpoints of dysfunctional telomeres. Investigating the enzymatic activity responsible for the random insertions enabled us to identify polymerase θ (Pol θ ; encoded by *Polq* in mice) as a crucial alternative NHEJ factor in mammalian cells. *Polq* inhibition suppresses alternative NHEJ at dysfunctional telomeres, and hinders chromosomal translocations at non-telomeric loci. In addition, we found that loss of *Polq* in mice results in increased rates of homology-directed repair, evident by recombination of dysfunctional telomeres and accumulation of RAD51 at double-stranded breaks. Lastly, we show that depletion of Pol θ has a synergistic effect on cell survival in the absence of *BRCA* genes, suggesting that the inhibition of this mutagenic polymerase represents a valid therapeutic avenue for tumours carrying mutations in homology-directed repair genes.

Chromosome end-to-end fusions are inhibited by shelterin; a multi-subunit complex anchored to telomeric DNA by two Myb-containing proteins—TRF1 and TRF2 (ref. 4). Telomere fusions are executed by two independent end-joining pathways. Classical non-homologous end-joining (C-NHEJ), mediated by LIG4 and the Ku70/80 heterodimer, is primarily blocked by TRF2 (ref. 5). Conversely, alternative NHEJ (alt-NHEJ), which is dependent on LIG3 (ref. 6) and PARP1 (ref. 7), is repressed in a redundant manner^{2,3}. Alt-NHEJ is fully unleashed after the simultaneous depletion of TRF1 and TRF2, and the creation of shelterin-free telomeres in cells deficient for *Ku70* and *Ku80* (also known as *Xrcc6* and *Xrcc5*, respectively)². This error-prone end-joining pathway mediates fusion of naturally eroded telomeres¹, joining of switch regions during class-switch recombination⁸, and formation of chromosomal translocations in mouse cells^{9,10}.

To characterize the differences between C-NHEJ and alt-NHEJ at dysfunctional telomeres, we determined whether the sequence of the junction between two fused telomeres differed depending on the type of repair pathway used. Telomere fusions by C-NHEJ were triggered by Cre-mediated depletion of TRF2 using previously described mouse embryonic fibroblasts (MEFs) (*Trf2*^{F/F}Cre-ER^{T2})¹¹ (Extended Data Fig. 1a). To induce robust fusions by the alt-NHEJ pathway, we depleted the entire shelterin complex by deleting *Trf1* and *Trf2* from *Trf1*^{F/F}*Trf2*^{F/F}*Ku80*^{−/−}Cre-ER^{T2} MEFs² (Extended Data Fig. 1a). DNA was subjected to next-generation sequencing, and reads corresponding to telomeres were identified on the basis of the presence of at least three consecutive TTAGGG repeats. To detect rare reads containing fusion junctions, we exploited the novel sequence arrangement created by the ligation of the 3' G-rich strand (TTAGGG-3') to the 5' C-rich strand (5'-CCCTAA) (Fig. 1a), and filtered reads that started with at least three G-rich repeats and ended with two or more C-rich repeats. We confirmed that this approach could

successfully identify telomere fusions by comparing reads derived from *Trf2*-proficient and *Trf2*-deficient cells. Starting with a similar number of telomere-repeat containing reads, we identified >90 fusogenic events in *Trf2*-knockout MEFs, compared to only three events in wild-type

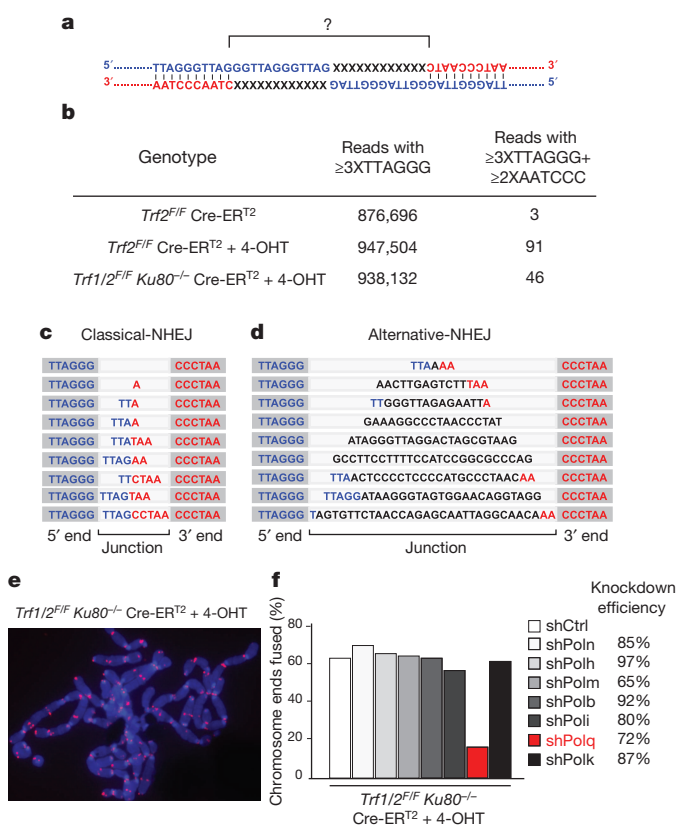


Figure 1 | Random nucleotide insertions at the junction of telomeres fused by alt-NHEJ. **a**, Schematic of the junction of a telomere fusion. The 3' end of the telomeric G-rich strand of a chromosome (blue) is fused to the 5' end of the C-rich strand of a different chromosome (red). **b**, Illumina sequencing to analyse telomere fusion junctions. Reads ≥ 3 XTTAGGG consecutively were scored as derived from telomere fragments. Those with ≥ 3 XTTAGGG on the 5' end and ≥ 2 XCCCTAA at the 3' end were scored as telomere fusion junctions (see Supplementary Information). **c**, Examples of telomere fusions generated by C-NHEJ of TRF2-depleted telomeres. Light grey highlights fusion junctions, dark grey marks the flanking telomere repeats. **d**, Examples of insertions in shelterin-free *Ku80*-null MEFs. **e**, Telomere fusions in metaphase spreads from *Trf1*^{F/F}*Trf2*^{F/F}*Ku80*^{−/−}Cre-ER^{T2} MEFs. Telomeres in red (peptide nucleic acid (PNA) probe) and chromosomes in blue (4',6-diamidino-2-phenylindole; DAPI). **f**, Frequency of telomere fusions after the depletion of candidate polymerases. 4-OHT, 4-hydroxytamoxifen; shCtrl, control shRNA. Bars represent mean of $n > 1,000$ chromosome ends derived from one experiment.

¹Skirball Institute of Biomolecular Medicine, Department of Cell Biology, NYU School of Medicine, New York, New York 10016, USA. ²Department of Molecular Biosciences, Institute for Cellular and Molecular Biology, University of Texas at Austin, 2506 Speedway Stop A5000, Austin, Texas 78712, USA. ³Department of Molecular and Experimental Medicine, The Scripps Research Institute, La Jolla, California 92037, USA.

cells (Fig. 1b). Sequence analysis of the junctions highlighted different permutations of TTAGGG/AATCCC sequences. Notably, the spectrum of the fusion junctions was different in shelterin-free settings, in which frequent non-telomeric nucleotide insertions (9 out of 46 events) were identified at fusion breakpoints (Fig. 1b–d and Supplementary Information).

To identify the enzyme that incorporated nucleotides at dysfunctional telomeres, we depleted known low-fidelity DNA polymerases in shelterin-free cells lacking *Ku80*, and analysed chromosome-end fusions on metaphase spreads. Notably, we observed a reduction in the frequency of telomere fusions in cells with reduced levels of polymerase theta (Polθ, encoded by *Polq* in mice) (Fig. 1e, f and Extended Data Fig. 1b). The activity of Polθ is specific to alt-NHEJ as its inhibition in *Trf2*-knockout cells did not affect the frequency of C-NHEJ (Fig. 2a, b and Extended Data Fig. 2a–c).

Polθ is an A-family DNA polymerase that exhibits low fidelity on templated DNA¹², and also displays a terminal transferase-like activity that catalyses nucleotide addition in a template-independent manner¹³. The relevance of these activities *in vivo* was highlighted in *Drosophila melanogaster*, in which Polθ was shown to stimulate nucleotide insertions during double-strand break (DSB) repair by alt-NHEJ¹⁴. More recently, Polθ was shown to promote end-joining of replication-associated DSBs in *Caenorhabditis elegans*¹⁵, preventing large deletions around G-rich DNA¹⁶. The exact function of Polθ during DSB repair in mammalian cells remains elusive.

The crucial role for Polθ at dysfunctional telomeres prompted us to test whether it is required for DSB repair at non-telomeric loci. To this end, we tested whether the depletion of Polθ affects chromosomal translocations in the context of mouse pluripotent cells, reported to be mediated by alt-NHEJ¹⁰ in a LIG3-dependent manner⁹. To model chromosomal translocations, we induced DSBs in the *Rosa26* and *H3f3b* mouse loci using the CRISPR/Cas9 system (Fig. 2c). When introduced into *Polq*^{+/+} and *Polq*^{-/-} cells¹⁷, the Cas9-gRNA(*Rosa26*; *H3f3b*) expression plasmid

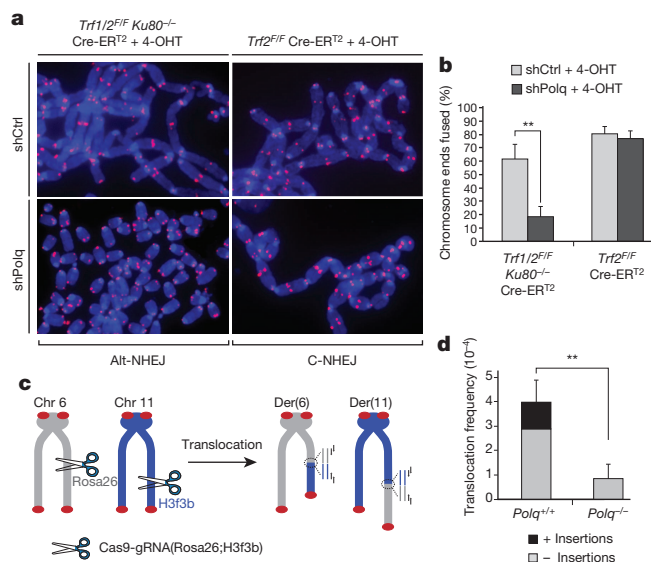


Figure 2 | Polθ is required for alt-NHEJ-dependent DSB repair in mammalian cells. **a**, Metaphases from *Trf2*^{2/2} Cre-ER^{T2} + 4-OHT and shelterin-free (*Trf1*^{2/2} *Trf2*^{2/2} *Ku80*^{-/-} Cre-ER^{T2} + 4-OHT) MEFs infected with the indicated short hairpin RNA (shRNA). **b**, Quantification of telomere fusions in MEFs with the indicated treatment (mean values ± s.d. derived from six independent experiments). ***P* = 0.003; two-tailed Student's *t*-test). **c**, Design of the translocation assay in which DSBs are induced by Cas9-gRNA(*Rosa26*; *H3f3b*). Joining of DNA ends generates der(6) and der(11), detected by nested PCR⁹. **d**, Translocation frequency in *Polq*^{+/+} and *Polq*^{-/-} cells 60 h after Cas9-gRNA(*Rosa26*; *H3f3b*) expression. Mean values ± s.d. derived from three independent experiments. ***P* = 0.009; two-tailed Student's *t*-test.

induced simultaneous cleavage of both loci with comparable efficiencies (Extended Data Fig. 2d, e). Consistent with previous reports, 24% of translocation events in *Polq*^{+/+} cells were scarred by random insertions. Interestingly, the overall frequency of translocations in cells lacking *Polq* was significantly reduced (Fig. 2d). Sequence analysis of residual translocations in *Polq*^{-/-} cells highlighted the absence of insertions, and a concomitant decrease in micro-homology at junctions (Fig. 2d and Extended Data Figs 2–5). Notably, we observed similar results when assessing translocation frequency in cells expressing a catalytically inactive form of Polθ (Extended Data Fig. 2g–k). Altogether, our data indicate that the promiscuous activity of Polθ during DSB repair contributes to the increased mutagenicity of alt-NHEJ. Importantly, our results indicate that mammalian Polθ plays a critical part by stimulating the end-joining reaction.

We next investigated the upstream signalling event(s) required for the recruitment of Polθ to DNA damage sites, induced after micro-irradiation of HeLa cells expressing Myc-tagged Polθ. Accumulation of Polθ at laser-induced DNA breaks, discerned by its co-localization with the phosphorylated histone H2AX (γ-H2AX), occurred in ~25% of cells that stained positive for Myc (Fig. 3a, b), and was independent of either ATM or ATR signalling (Extended Data Fig. 6). Instead,

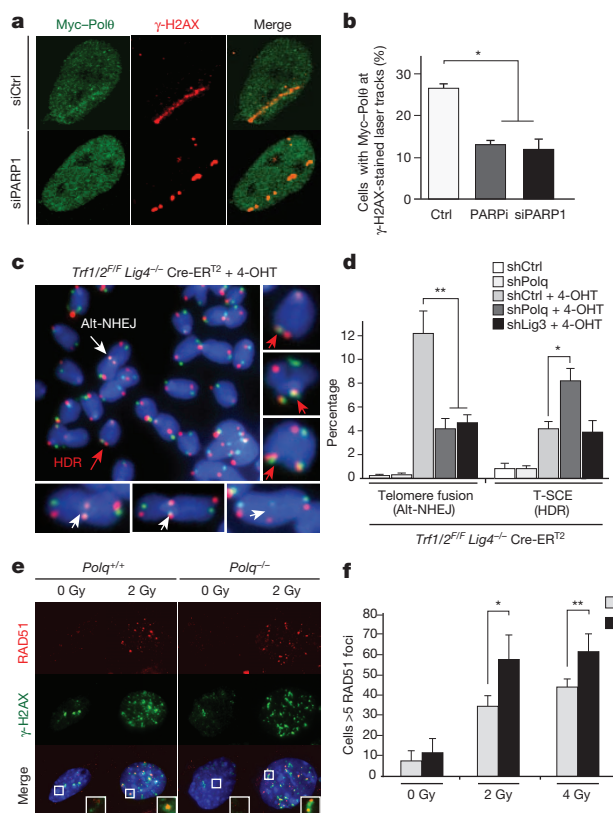


Figure 3 | Polθ is recruited by PARP1 to promote alt-NHEJ at the expense of HDR. **a**, Myc-Polθ localization to DNA damage was monitored after laser micro-irradiation of HeLa cells. Cells were fixed and stained for γ-H2AX and Myc, 1 h after damage induction. **b**, Quantification of Polθ accumulation at sites of laser damage (mean values ± s.e.m. derived from two independent experiments). **c**, To test whether Polθ represses recombination at telomeres, we depleted the polymerase in shelterin-free and *Lig4*-deficient MEFs², and both repair pathways were monitored using CO-FISH. White arrows indicate alt-NHEJ events, red arrows highlight HDR-mediated T-SCEs. **d**, Quantification of telomere fusion (alt-NHEJ) and T-SCE (HDR) in cells transduced with shRNAs against *Polq*, *Lig3* or control shRNA. Error bars denote ± s.d. from three independent experiments. **e**, Immunofluorescence for RAD51 and γ-H2AX in the indicated MEFs 3 h after irradiation. **f**, Graph representing quantification of ionizing-radiation-induced RAD51 foci. Mean values ± s.d. derived from three independent experiments. **P* < 0.05, ***P* < 0.01; two-tailed Student's *t*-test.

co-localization of Polθ with γ-H2AX was reduced after depletion of PARP1 with short interfering RNAs (siRNAs), or after the inhibition of PARP1 activity (using KU58948) (Fig. 3a, b). In a parallel experiment that used a recently developed U2OS-DSB reporter cell line¹⁸, we were able to ascertain the localization of Myc-Polθ to bona fide DSBs, induced after FOK1 cleavage of a LacO-tagged genomic locus (Extended Data Fig. 7). In conclusion, our data suggest that PARP1, previously known to be required for alt-NHEJ^{7,19}, facilitates the recruitment of Polθ to DSBs.

Homology-directed repair (HDR) is prevalent during the S/G2 phase of the cell cycle, which coincides with the peak of alt-NHEJ activity, and these pathways also share the initial resection step mediated by MRE11 and CtIP²⁰. To test whether inhibiting alt-NHEJ could potentially result in increased HDR, we depleted shelterin in *Lig4*-deficient MEFs, a genetic setting that is conducive to the activity of alt-NHEJ as well as HDR². To investigate the relative contribution of the two repair pathways, we used a chromosome-orientation fluorescence *in situ* hybridization (CO-FISH) assay²¹, and monitored the exchange of telomeres between sister chromatids by HDR (telomere sister chromatid exchange, T-SCE), and, at the same time, measured the frequency of chromosome end-end fusion by end-joining (Fig. 3c). After depletion of shelterin from *Trf1*^{F/F} *Trf2*^{F/F} *Lig4*^{-/-} Cre-ER^{T2} MEFs, ~10% of the telomeres were processed by alt-NHEJ, whereas ~5% of chromosome ends showed T-SCEs² (Fig. 3c, d and Extended Data Fig. 8a–c). As expected, we observed a substantial reduction in the frequency of alt-NHEJ at shelterin-free telomeres in *Lig4*^{-/-} cells that lack *Polq* or *Lig3* (Fig. 3d and Extended Data Fig. 8a–e). Remarkably, *Polq*-depleted cells exhibited a concomitant increase in T-SCE, which was not evident in cells lacking *Lig3* (Fig. 3d), thereby highlighting a unique role for Polθ in counteracting HDR. To gain insight into this novel Polθ function, we show that the promiscuous polymerase is not required for end-resection of DSBs (Extended Data Fig. 8f, g). Instead, its activity counteracts the accumulation of RAD51 foci (Fig. 3e, f and Extended Data Fig. 8h). To corroborate these findings, we used the traffic light reporter (TLR) system, designed to generate a flow-cytometric readout for HDR and end-joining at a site-specific DNA break induced by I-Sce1 (ref. 22). We observed that after knocking down *Polq* in *Lig4*^{-/-} cells, resolution of the I-Sce1-induced DNA break by HDR is increased, in conjunction with a significant reduction in the frequency of alt-NHEJ (Extended Data Fig. 9).

Alt-NHEJ is often considered as a back-up choice for DSB repair, operating at the expense of genomic stability. Circumstantial evidence suggests that this pathway could be enhanced when HDR is impaired^{23,24}. We therefore postulated that this error-prone mode of repair has an essential role in cells with compromised HDR activity. We tested this hypothesis by inhibiting *Polq* in cells lacking the breast cancer susceptibility genes—*Brca1* and *Brca2*. Chromosome analysis revealed a four-fold increase in chromosomal aberrancies after *Polq* depletion in MEFs lacking either *Brca1* or *Brca2*. Such aberrancies included chromatid and chromosome breaks, in addition to radial chromosome structures characteristic of *Lig4*-mediated processing of chromatid breaks via the C-NHEJ pathway (Fig. 4a, b and Extended Data Fig. 10a, b). Ultimately, the increased genomic instability in cells co-depleted for *Polq* and *Brca* genes compromised cellular survival. We observed that *BRCA1*-mutated human cells (Fig. 4c, d), and mouse cells lacking *Brca1* (Extended Data Fig. 10c–f), displayed significantly reduced colony-forming capabilities after *Polq* impairment. Although we cannot exclude that Polθ performs additional activities required for the survival of *Brca*-deficient cells²⁵, our data suggest that Polθ-mediated alt-NHEJ promotes the survival of cells with a compromised HDR pathway. In the absence of a safer means to repair breaks, alt-NHEJ may therefore prevent genomic havoc by resolving unrepaired lesions.

Here we provide direct evidence linking Polθ to alt-NHEJ repair in mammalian cells (Fig. 4e). We also show that while Polθ hinders error-free repair by HDR, its activity is essential for the survival of HDR-deficient cells (Fig. 4e). The question remains as to how this promiscuous polymerase orchestrates DSB repair. After DSB formation in the S/G2 phase

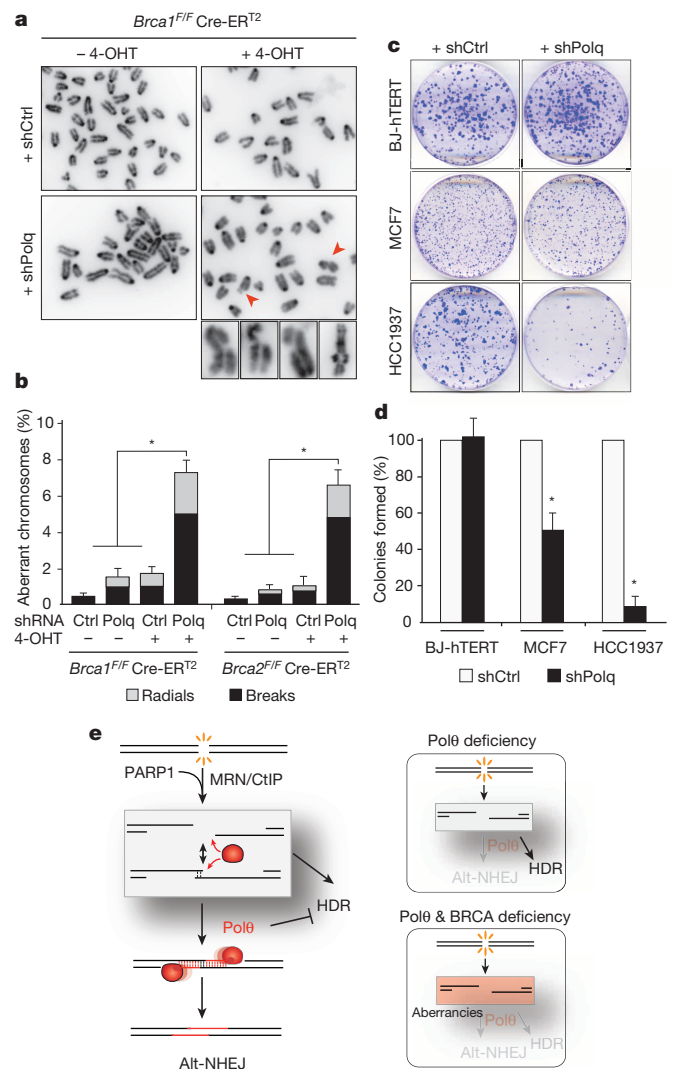


Figure 4 | *Polq* inhibition in *Brca*-mutant cells leads to increased chromosomal aberrancies and reduced cellular survival. **a**, Analysis of genomic instability in metaphase spreads from *Brca1*^{F/F} Cre-ER^{T2} MEFs treated with shRNA against *Polq* or vector control. **b**, Quantification of breaks (chromatid and chromosome) and radials in *Brca1*^{F/F} Cre-ER^{T2} and *Brca2*^{F/F} Cre-ER^{T2} MEFs with the indicated treatment. Mean values are presented with error bars denoting ± s.d. from three independent experiments. **c**, Clonogenic survival after *Polq* depletion. Crystal violet staining of BJ-hTERT, MCF7 and HCC1937 cells treated with shRNA against *Polq* or vector control. **d**, Quantitative analyses of colony formation assay. Colonies in each control shRNA cell line were set to 100%. Colonies in shPolq-expressing cells are normalized to shCtrl. Mean values ± s.d. derived from three independent experiments. **P* < 0.05, ***P* < 0.01; two-tailed Student's *t*-test. **e**, Schematic depicting our model for the function of Polθ during DSB repair (see Supplementary Information).

of the cell cycle, resection of DSBs by MRE11 and CtIP²⁰ potentially exposes micro-homology that allows spontaneous annealing of broken DNA ends (Fig. 4e). The binding of RPA antagonizes this annealing step to promote HDR-mediated repair^{26,27}. An opposing activity is likely to be exerted by Polθ. Placing our finding in the context of recent biochemical experiments and genetic studies in model organisms^{13–16}, we ultimately propose a model in which Polθ exploits both its template-independent and template-dependent activities to stabilize the annealed intermediate structure and channel repair towards the alt-NHEJ pathway (Fig. 4e). Our findings that Polθ is critical for alt-NHEJ support this model, which provides a potential explanation as to how this polymerase counteracts HDR.

Finally, it is intriguing that although *POLQ* expression in normal human tissues is generally repressed²⁸, it is upregulated in a wide range of human cancers and associates with poor clinical outcome in breast tumours^{29,30}. Our findings that cells with compromised HDR activity depend on this mutagenic polymerase for survival establish a rationale for the development of Polθ-targeted approaches for cancer treatment.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 28 July; accepted 16 December 2014.

Published online 2 February 2015.

- Capper, R. *et al.* The nature of telomere fusion and a definition of the critical telomere length in human cells. *Genes Dev.* **21**, 2495–2508 (2007).
- Sfeir, A. & de Lange, T. Removal of shelterin reveals the telomere end-protection problem. *Science* **336**, 593–597 (2012).
- Rai, R. *et al.* The function of classical and alternative non-homologous end-joining pathways in the fusion of dysfunctional telomeres. *EMBO J.* **29**, 2598–2610 (2010).
- Sfeir, A. Telomeres at a glance. *J. Cell Sci.* **125**, 4173–4178 (2012).
- van Steensel, B., Smogorzewska, A. & de Lange, T. TRF2 protects human telomeres from end-to-end fusions. *Cell* **92**, 401–413 (1998).
- Wang, H. *et al.* DNA ligase III as a candidate component of backup pathways of nonhomologous end joining. *Cancer Res.* **65**, 4020–4030 (2005).
- Audebert, M., Salles, B., Weinfeld, M. & Calsou, P. Involvement of polynucleotide kinase in a poly(ADP-ribose) polymerase-1-dependent DNA double-strand breaks rejoining pathway. *J. Mol. Biol.* **356**, 257–265 (2006).
- Yan, C. T. *et al.* IgH class switching and translocations use a robust non-classical end-joining pathway. *Nature* **449**, 478–482 (2007).
- Simsek, D. *et al.* DNA ligase III promotes alternative nonhomologous end-joining during chromosomal translocation formation. *PLoS Genet.* **7**, e1002080 (2011).
- Simsek, D. & Jasin, M. Alternative end-joining is suppressed by the canonical NHEJ component Xrcc4-ligase IV during chromosomal translocation formation. *Nature Struct. Mol. Biol.* **17**, 410–416 (2010).
- Okamoto, K. *et al.* A two-step mechanism for TRF2-mediated chromosome-end protection. *Nature* **494**, 502–505 (2013).
- Arana, M. E., Seki, M., Wood, R. D., Rogozin, I. B. & Kunkel, T. A. Low-fidelity DNA synthesis by human DNA polymerase theta. *Nucleic Acids Res.* **36**, 3847–3856 (2008).
- Hogg, M., Sauer-Eriksson, A. E. & Johansson, E. Promiscuous DNA synthesis by human DNA polymerase θ. *Nucleic Acids Res.* **40**, 2611–2622 (2012).
- Chan, S. H., Yu, A. M. & McVey, M. Dual roles for DNA polymerase theta in alternative end-joining repair of double-strand breaks in *Drosophila*. *PLoS Genet.* **6**, e1001005 (2010).
- Roerink, S. F., van Schendel, R. & Tijsterman, M. Polymerase theta-mediated end joining of replication-associated DNA breaks in *C. elegans*. *Genome Res.* **24**, 954–962 (2014).
- Koole, W. *et al.* A Polymerase Theta-dependent repair pathway suppresses extensive genomic instability at endogenous G4 DNA sites. *Nature Commun.* **5**, 3216 (2014).
- Shima, N. *et al.* Phenotype-based identification of mouse chromosome instability mutants. *Genetics* **163**, 1031–1040 (2003).
- Tang, J. *et al.* Acetylation limits 53BP1 association with damaged chromatin to promote homologous recombination. *Nature Struct. Mol. Biol.* **20**, 317–325 (2013).
- Wang, M. *et al.* PARP-1 and Ku compete for repair of DNA double strand breaks by distinct NHEJ pathways. *Nucleic Acids Res.* **34**, 6170–6182 (2006).
- Truong, L. N. *et al.* Microhomology-mediated End Joining and Homologous Recombination share the initial end resection step to repair DNA double-strand breaks in mammalian cells. *Proc. Natl Acad. Sci. USA* **110**, 7720–7725 (2013).
- Bailey, S. M., Cornforth, M. N., Kurimasa, A., Chen, D. J. & Goodwin, E. H. Strand-specific postreplicative processing of mammalian telomeres. *Science* **293**, 2462–2465 (2001).
- Certo, M. T. *et al.* Tracking genome engineering outcome at individual DNA breakpoints. *Nature Methods* **8**, 671–676 (2011).
- Bindra, R. S., Goglia, A. G., Jasin, M. & Powell, S. N. Development of an assay to measure mutagenic non-homologous end-joining repair activity in mammalian cells. *Nucleic Acids Res.* **41**, e115 (2013).
- Sakai, W. *et al.* Secondary mutations as a mechanism of cisplatin resistance in BRCA2-mutated cancers. *Nature* **451**, 1116–1120 (2008).
- Fernandez-Vidal, A. *et al.* A role for DNA polymerase θ in the timing of DNA replication. *Nature Commun.* **5**, 4285 (2014).
- Deng, S. K., Gibb, B., de Almeida, M. J., Greene, E. C. & Symington, L. S. RPA antagonizes microhomology-mediated repair of DNA double-strand breaks. *Nature Struct. Mol. Biol.* **21**, 405–412 (2014).
- Chen, H., Lisby, M. & Symington, L. S. RPA coordinates DNA end resection and prevents formation of DNA hairpins. *Mol. Cell* **50**, 589–600 (2013).
- Kawamura, K. *et al.* DNA polymerase theta is preferentially expressed in lymphoid tissues and upregulated in human cancers. *Int. J. Cancer* **109**, 9–16 (2004).
- Higgins, G. S. *et al.* Overexpression of POLQ confers a poor prognosis in early breast cancer patients. *Oncotarget* **1**, 175–184 (2010).
- Lemée, F. *et al.* DNA polymerase theta up-regulation is associated with poor survival in breast cancer, perturbs DNA replication, and promotes genetic instability. *Proc. Natl Acad. Sci. USA* **107**, 13390–13395 (2010).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank T. de Lange, R. Greenberg, J. Shay, N. Shima, C. Cazaux and R. Wood for providing key reagents for this study. We are grateful to M. Ji, L. Walton Masters, A. Phillips, A. Pinzaru, F. Yeung, P. Tonzi and J. Wong for technical assistance. We thank S. Kabir and F. Lottersberger for critical reading of the manuscript. This work was supported by a grant from the Breast Cancer Alliance (A.S.), V-foundation (A.S.), Department of Defense Breast Cancer Research Program BC134020 (P.A.M.-G.), Pew-Stewart Scholars Award (A.S.), Pew Scholars Award (E.L.-D.), Novartis Advanced Discovery Institute (E.L.-D.), and a grant from the National Institutes of Health (NIH) AG038677 (E.L.-D.). The A.S. laboratory was supported by start-up funds from the Helen L. and Martin S. Kimmel Center for Stem Cell Biology. The K.M.M. laboratory was supported in part by start-up funds from the University of Texas at Austin and from the Cancer Prevention Research Institute of Texas (CPRIT, R116). K.M.M. is a CPRIT scholar.

Author Contributions A.S., E.L.-D. and P.A.M.-G. conceived the experimental design. P.A.M.-G. and A.S. performed the experiments and analysed the data. E.L.-D. and N.N. performed telomere-sequencing experiments. F.G. and K.M.M. performed experiments related to Polθ localization at DNA breaks. A.S. wrote the manuscript. All authors discussed the results and commented on the manuscript.

Author Information Sequence has been deposited with the BioProject database under accession PRJNA269507. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.S. (agnel.sfeir@med.nyu.edu).

METHODS

Cell culture procedures. *Trf2^{F/F}* Cre-ER^{T2}, *Trf1^{F/F}* *Trf2^{F/F}* *Ku80^{-/-}* Cre-ER^{T2} and *Trf1^{F/F}* *Trf2^{F/F}* *Lig4^{-/-}* Cre-ER^{T2} MEF lines were previously described²¹. *Polq^{+/-}* and *Polq^{-/-}* MEFs were a gift from N. Shima¹⁷. *Brca1^{F/F}* Cre-ER^{T2} and *Brca2^{F/F}* Cre-ER^{T2} MEFs and U2OS-DSB reporter cells¹⁸ were a gift from R. Greenberg. *Trf1^{F/F}* *Trf2^{F/F}* *Lig4^{-/-}* Cre-ER^{T2} MEFs were derived from mice that were deficient in p53. The remaining MEF lines were immortalized with pBabeSV40LargeT. MEFs were cultured in DMEM supplemented with 10–15% FBS (Gibco), 2 mM L-glutamine (Sigma), 100 U ml⁻¹ penicillin (Sigma), 0.1 µg ml⁻¹ streptomycin (Sigma), 0.1 mM non-essential amino acids (Invitrogen) and 1 mM sodium pyruvate (Sigma). Expression of Cre recombinase was induced by treating MEFs carrying the Cre-ER^{T2} allele with 0.5 µM 4-OHT (Sigma H7904) for 12 h. The *t* = 0 time point was set at the time of treatment with 4-OHT. BJ-hTERT and MCF7 cells were grown in DMEM supplemented with 10% FBS. HCC1937 cells were grown in RPMI medium (Gibco) containing 15% FBS. U2OS-DSB reporter cells were grown in DMEM supplemented with 10% BCS. Human HeLa cells were grown in DMEM supplemented with 10% FBS, 100 U ml⁻¹ penicillin, 100 µg ml⁻¹ streptomycin and 2 mM L-glutamine. Mouse embryonic stem cells were grown in DMEM supplemented with 15% FBS (ES-qualified FBS) (Gibco), 2 mM L-glutamine (Sigma), 100 U ml⁻¹ penicillin (Sigma), 0.1 µg ml⁻¹ streptomycin (Sigma), 0.1 mM non-essential amino acids (Invitrogen), leukaemia inhibitory factor (LIF) and 2-β-mercaptoethanol (Gibco 21985). For inhibitor experiments, PARPi (KU58948, Axon medchem), ATMi (KU-55933, Tocris) and ATRi (VE-821, Selleckchem) were all used at a final concentration of 10 µM, and were applied to culture medium 2–4 h before irradiation. For ionizing radiation treatment, cells were exposed to 1–10 Gy ionizing radiation by a Faxitron X-ray system (120 kV, 5 mA, dose rate 5 Gy min⁻¹) and recovered for 4 h before immunofluorescence analysis.

Lentiviral delivery of shRNA. shRNA treatments were carried out before 4-OHT treatment. shRNAs (see below for a list of sequences) were introduced by two lentiviral infections at 12 h intervals using supernatant from transfected 293T cells. Parallel infection with pLKO.1 was used as a negative control. Cells were selected with puromycin for 3 days.

TLR assay. Lentiviral constructs coding for TLR (31482) and I-SceI with donor e-GFP (31476) were purchased from Addgene²². To avoid the confounding effect of classical-NHEJ on the repair of I-SceI-induced DNA breaks, we stably integrated the TLR construct into *Ku80^{-/-}* and *Lig4^{-/-}* MEFs. The plasmid was transduced by two lentiviral infections at 12 h intervals using supernatant from transfected 293T cells. Cells with integrated TLR were selected with puromycin for 5 days. Cells were then transduced with concentrated *Polq* shRNA lentiviral particles followed by I-SceI. Cells were collected 72 h later without further antibiotic selection and analysed on a BD LSRII. eGFP fluorescence, which reflects HDR repair, was measured using a 488-nm laser for excitation and a 530/30 filter for detection. mCherry fluorescence, indicative of alt-NHEJ was measured by using a 561-nm laser for excitation and a 610/20 filter for detection. Data were analysed using FlowJo software.

Detection of telomeric fusions. To enrich for telomeric DNA, genomic DNA was digested with two frequent cutters (AluI and MboI) and fragments greater than 10 kilobases (kb) were isolated. The resulting DNA was used to generate a library using the NEBNext Ultra Library Prep Kit and the NEBNext Multiplex Oligos for Illumina (NEB) following the manufacturer instructions. The resulting library was run on an Illumina HiSeq platform generating 100-base-pair (bp) indexed pair-end reads.

Transient transfection of cells and laser micro-irradiation. Full-length human *POLQ* was cloned into pLPC-Myc vector. HeLa cells were plated on glass-bottomed dishes (Willco Wells). Myc-Polθ constructs were transfected into HeLa and U2OS-DSB reporter cells cell with HilyMax (Dojindo) according to the manufacturer's instruction. Then, 24 h after transfection, cells were pre-sensitized with 10 µM 5-bromo-2'-deoxyuridine (BrdU) in normal DMEM medium for 20 h. After indicated treatments, cells were damaged by laser micro-irradiation as previously described³¹. After laser micro-irradiation, cells were incubated for 1 h, then fixed and analysed by immunofluorescence and microscopic imaging as described below. For PARP1 siRNA experiments, cells were transfected with siCtrl (non-targeting pool, Thermo Scientific) or siPARP1 (GGGCAAGCACAGUGUCAAUU, Sigma), for 24 h before *POLQ* transfections and subsequent treatments as described above.

Immunofluorescence and confocal microscopy. After the indicated treatments, cells were processed and analysed for immunofluorescence as previously described³². In brief, cells were fixed with 2% (v/v) paraformaldehyde for 15 min at room temperature. Cells were washed with PBS, permeabilized with 0.5% (v/v) Triton X-100 for 10 min, and blocked with PBS containing 3% BSA. Cells were incubated with the same buffer containing primary antibodies for 1 h at room temperature followed by secondary antibodies incubations for 1 h at room temp. Cells were imaged and analysed with Z-stacked setting using the FV10-ASW3.1 software on a Fluoview 1000 confocal microscope (Olympus). For laser line quantification, >50 cells were counted for all conditions from two independent experiments. The primary

antibodies used for immunofluorescence were γ-H2AX (p Ser139) (rabbit polyclonal, Novus, NB100-384) and c-Myc (mouse monoclonal, Santa Cruz, sc-40). The secondary antibodies used for immunofluorescence were Alexa Fluor 594 (rabbit) (Invitrogen, A11037) and Alexa Fluor 488 (mouse) (Invitrogen, A11029). To analyse the recruitment of Polθ to double-stranded breaks, U2OS-DSB reporter cells expressing Myc-Polθ were analysed 4 h after treatment with shield and tamoxifen. Lastly, to analyse RAD51 foci formation and its co-localization with γ-H2AX (p Ser139) after ionizing radiation treatment, cells were treated with 0.2% Triton X-100 (in PBS) for 5 min on ice before fixation with paraformaldehyde. The primary antibodies used for RAD51 immunofluorescence were γ-H2AX (p Ser139) (mouse monoclonal, Novus, NB100-384) and RAD51 (rabbit polyclonal, Santa Cruz, sc-8349).

FISH. Cells were collected at 96 h after treatment with 4-OHT to analyse the frequency of telomere fusions. In brief, ~80% confluent MEFs were incubated for 2 h with 0.2 µg ml⁻¹ colcemid (Sigma). The cells were collected by trypsinization, resuspended in 0.075 M KCl at 37 °C for 30 min, and fixed overnight in methanol/acetic acid (3:1) at 4 °C. The cells were dropped onto glass slides and the slides were dried overnight. The next day, the slides were rehydrated with PBS for 15 min then fixed with 4% formaldehyde for 2 min at room temperature. Slides were digested with 1 mg ml⁻¹ pepsin, pH 2.2, at 37 °C for 10 min, washed three times with PBS and fixed again in 4% formaldehyde for 2 min at room temperature. After three PBS washes, the slides were incubated consecutively with 75%, 95% and 100% ethanol and allowed to air dry for 30 min before applying hybridization solutions (70% formamide, 1 mg ml⁻¹ blocking reagent (Roche), 10 mM Tris-HCl, pH 7.2) containing TAMRA-OO-(TTAGGG)₃ PNA probes (Applied Biosystems). Slides were denatured by heating for 3 min at 80 °C and hybridized for 2 h at room temperature. The next day, the slides were washed twice for 15 min each in 70% formamide, 10 mM Tris-HCl, followed by three 5-min washes in 0.1 M Tris-HCl, pH 7.0, 0.15 M NaCl and 0.08% Tween-20. Chromosomal DNA was counterstained with DAPI during the second PBS wash. Slides were mounted in antifade reagent (Pro-Long Gold, Invitrogen) and images were captured with a Nikon Eclipse TI microscope (see <http://delangelab.rockefeller.edu/protocols>).

CO-FISH. Cells were labelled with BrdU:BrdC (3:1, final concentration 10 µM) for 14–16 h. Two hours before collection by trypsinization, 0.2 µg ml⁻¹ colcemid was added to the media. To fix the cells and drop metaphases on a glass slide, the same procedure that was applied for FISH was followed. Slides were treated with 0.5 mg ml⁻¹ RNase A (in PBS, DNase-free) for 10 min at 37 °C, incubated with 0.5 µg ml⁻¹ Hoechst 33258 (Sigma) in 2×SSC for 15 min at room temperature, and exposed to 365-nm ultraviolet light (Stratalinker 1800 UV irradiator) for 30 min. The slides were then digested twice with 800 U exonuclease III (Promega) at room temperature for 10 min each, washed with PBS and dehydrated through an ethanol series of 70%, 95% and 100%. After air-drying, slides were hybridized with Tamra-OO-(TTAGGG)₃ PNA probe in hybridization solution (70% formamide, 1 mg ml⁻¹ blocking reagent (Roche) and 10 mM Tris-HCl, pH 7.2) for 2 h at room temperature. The slides were then washed for a few seconds with 70% formamide and 10 mM Tris-HCl, pH 7.2, and incubated with FITC-OO-(CCCTAA)₃ PNA probe in hybridization solution for 2 h. Slides were washed and mounted as described for FISH (see <http://delangelab.rockefeller.edu/protocols>).

Western blot analysis. Cells were collected by trypsinization, lysed in 2× Laemmli buffer (100 mM Tris-HCl, pH 6.8, 200 µM dithiothreitol, 3% SDS, 20% glycerol and 0.05% bromophenol blue) at 1 × 10⁴ cells per microlitre. The lysate was denatured for 10 min at 95 °C, and sheared by forcing it through a 28-gauge insulin needle ten times. Lysate from 1 × 10⁵ cells was loaded on an SDS-PAGE and transferred to a nitrocellulose membrane. The membrane was blocked in 5% milk in TBS with 0.1% Tween-20 and incubated with primary antibody in TBS, 5% milk and 0.1% Tween-20 for 2 h at room temperature. The following primary antibodies were used: Polθ (ab80906, Abcam); TRF1 (1449, rabbit polyclonal); RAP1 (1252, rabbit polyclonal); phospho-CHK2 (Thr68) (rabbit polyclonal, Cell Signaling); CHK2 (rabbit polyclonal, Cell Signaling); phospho-CHK1 (Ser 345) (mouse monoclonal, Cell Signaling); CHK1 (mouse monoclonal, Santa Cruz); LIG3 (mouse monoclonal, Santa Cruz); Myc (9E10; Calbiochem); and γ-tubulin (clone GTU-88, Sigma); PARP1 (polyclonal, Cell signaling). (See <http://delangelab.rockefeller.edu/protocols>.)

Chromosomal aberrancies. Cells were collected and dropped on microscope slides as described for the FISH protocol. After the slides had dried overnight, they were rehydrated in PBS, stained with 0.25 µg ml⁻¹ DAPI, dehydrated in a 70%, 95% and 100% ethanol series, mounted and imaged using Nikon Eclipse TI microscope. Aberrancies were scored as a percentage of chromatid breaks, chromosome breaks, and chromosome radial structures compared to total number of chromosomes.

Chromosomal translocation assay. Induced pluripotent stem cells were derived from primary *Polq^{+/-}* and *Polq^{-/-}* MEFs according to standard Yamanaka protocol³³. To perform the translocation assay, *Polq^{+/-}* and *Polq^{-/-}* induced pluripotent stem cells were transfected with 2 µg of Cas9-gRNA (Rosa26;H3f3b) plasmid per million cell. We constructed Cas9-gRNA (Rosa26;H3f3b) by introducing two guide RNAs

(5'-GTTGGCTCGCCGGATACGGG-3' for H3f3b; 5'-ACTCCAGTCTTTCTA GAAGA-3' for Rosa26) into pX330 (Addgene, 42230). After transfection, 1×10^4 cells were seeded per well in a 96-well plate, and lysed 3 days later in 40 μ l lysis buffer (10 mM Tris, pH 8.0, 0.45% Nonidet P-40 and 0.45% Tween 20). The lysate was incubated with 200 μ g ml⁻¹ proteinase K for 2 h at 55 °C. Translocation detection was performed according to previously established protocol⁹, using nested PCR. The primers used for the first PCR reaction were Tr6-11-Fwd: 5'-GCGGG AGAAATGGATATGAA-3' and Tr6-11-Rev: 5'-TTGACGCCTTCCTTCTTCT G-3' for der(6), and Tr11-6-Fwd: 5'-AACCTTTGAAAAAGCCACACA-3' and Tr11-6-Rev: 5'-GCACGTTTCCGACTTGAGTT-3', for der(11). For the second round of PCR amplification we used the primers Tr6-11NFwd: 5'-GGCGGAT CACAAGCAATAAT-3' and Tr6-11NRev: 5'-CTGCCATTCCAGAGATTGGT-3', and Tr11-6NFwd: 5'-AGCCACAGTGCTCACATCAC-3' and Tr11-6NRev: 5'-TCCCAAAGTCGCTCTGAGTT-3'. The number of PCR-positive wells was used to calculate the translocation frequency as previously described⁹. Amplified products from positive wells were sequenced to verify translocations and determine the junction sequences.

Surveyor assay. Forty-eight hours after transfection, genomic DNA was extracted with GE Healthcare Illustra Genomic Prep Mini Spin Kit (28-9042-76). The genomic region encompassing the guide RNA target sites was amplified using Q5 High-Fidelity DNA polymerase (New England BioLabs) with the primers Rosa26-Fwd: 5'-TAAAACTCGGGTGAGCATGT-3' and Rosa26-Rev: 5'-GGAGTTCTCTGC TGCTCCTG-3', and H3f3b-Fwd: 5'-GCGGCGGCTTGATTGCTCCAG-3' and H3f3b-Rev: 5'-AGCAACTGTCTACTCTGAGCCAC-3'. PCR fragments were gel purified and the surveyor assay was performed using a detection kit (Transgenomic), according to manufacturer's instructions. Agarose gels (2%) were used to visualize the bands after surveyor digestion.

Colony formation assay. After lentiviral transduction with shCtrl or (sequences listed below), cells were selected with puromycin (BJ: 0.5 μ g ml⁻¹; MCF7 and HCC1937: 1 μ g ml⁻¹; MEFs: 2 μ g ml⁻¹) for 72 h and plated in 6-cm dishes (1,000 and 10,000 cells per plate). After 10–14 days, colonies were fixed with 3% paraformaldehyde (5 min), rinsed with PBS, and stained with crystal violet (Sigma-Aldrich).

CRISPR targeting to mutate Polq gene in mouse embryonic stem cells. To generate cells carrying a catalytic dead Polq, two mutations at residues Asp2494Gly and Glu2495Ser (ref. 34) were introduced in the endogenous Polq locus in mouse embryonic stem cells using CRISPR/Cas9 gene targeting. Two guides RNAs were co-transfected with a Cas9-nickase (pX335-U6-Chimeric_BB-CBh-hSpCas9n(D10A)), and a donor cassette that introduces a SacII restriction site while replacing the two amino acid residues. Clonal cell lines were derived and genotyped to determine successful targeting. Two independent clonally derived lines were used for the analysis of translocation.

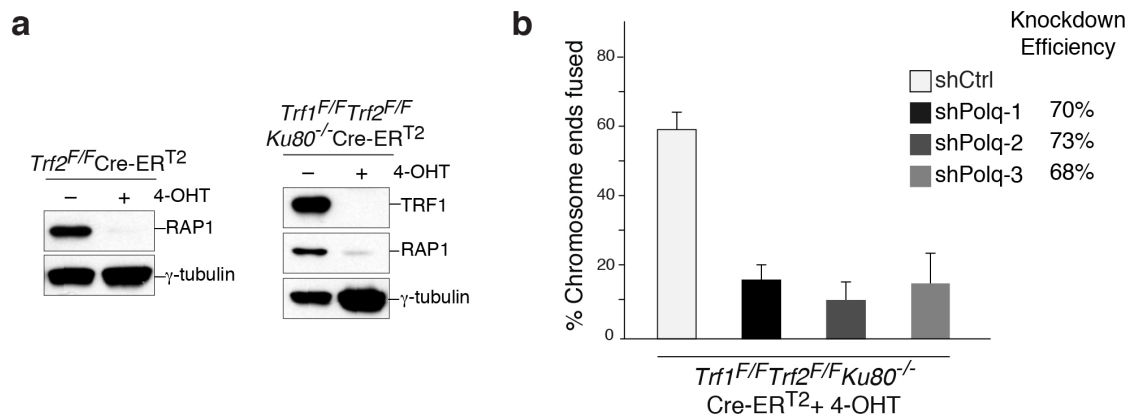
shRNA target sequence (pLK0.1 vector). shPolm: 5'-CTCACCTCTCACACAC CATAA-3'; shPolk: 5'-GCCCTTAGAAATGTCTCATAA-3'; shPolh: 5'-GCTC GATTCTCCAGCTTACAA-3'; shPoli: 5'-AGTGAAGAAGATACGTTTAAA-3'; shPolb: 5'-CCAAAGTTGTACATCGTGTT-3'; shPoln: 5'-CCTACTCACAT

GAAGGACATT-3'; shLig3 (mouse): 5'-CCAGACTTCAAACGTCTCAAA-3'; shLIG3 (human): 5'-CCGGATCATGTTCTCAGAAAT-3'; shPolq-1 (mouse): 5'-CGGCGGAGTATGAGAACTATT-3'; shPolq-2(mouse): 5'-CCAGGAATCA AAGACGACAAT-3'; shPolq-3(mouse): 5'-CCTGGCTGAATGCTGAACCTTT-3'; shPOLQ (human): 5'-CGGGCCTCTTTAGATATAAAT-3'; shBrca1(mouse): 5'-G CTCAGTGTATGACTCAGTTT-3'.

Primers for quantitative PCR. BRCA1Fwd: 5'-CTGCCGTCCAAATTCAGA AGT-3' and BRCA1Rev: 5'-CTGTGCTTCCCTGTAGGCT-3'; BRCA2Fwd: 5'-T GTGGTAGATGTTGCTAGTCCGCC-3' and BRCA2Rev: 5'-GCTTTTCTCGT TGTAAGTACTGCC-3'; POLbFwd: 5'-TGAACCATCATCAACGAATTGGG-3' and POLbRev: 5'-CCATGTCTCCACTCGACTCTG-3'; POLmFwd: 5'-AGGCT TCCGCTCCTAGAT-3' and POLmRev: 5'-GTGGGGAGAGCATCCATGTT-3'; POLkFwd: 5'-AGCTCAAATTACCAGCCAGCA-3' and POLkRev: 5'-GGTTG TCCCTCATTTCCACAG-3'; POLhFwd: 5'-ATCGAGTGGTTGCTCTGTAG A-3' and POLhRev: 5'-CCAAATGCTCGGGCTTCATAG-3'; POLiFwd: 5'-GC AGTCAAGGGCCACCTAC-3' and POLiRev: 5'-AGGTCTGTCTTTAATTCT GGGT-3'; POLnFwd: 5'-AGCTGATGGATGCTCTCAAGCAGG-3' and POLnRev: 5'-GAGTCAGAGTGCTGTTGCCTACATGG-3'; LIG3(mouse)Fwd: 5'-GAAG AAAGCTGCTGTCCAGG-3' and LIG3(mouse)Rev: 5'-CAGAGTTGTTGGGTT TTGCTG-3'; LIG3(human)Fwd: 5'-GAAGAGCTGGAAGATAATGAGAAGG-3' and LIG3(human)Rev: 5'-AGTGGTTGTCAACTAGCCTGG-3'; POLQ(mouse)Fwd: 5'-CAAGGTTTCATTCGGGTCTTGG-3' and POLQ (mouse)Rev: 5'-CGAGC AGGAAGATTCATCCAG-3'; POLQ(human) Fwd: 5'-CAGCCCTTATAGTG GAAGAAGC-3' and POLQ(human) Rev: 5'-GCACATGGATTCCATTGCAC TC-3'.

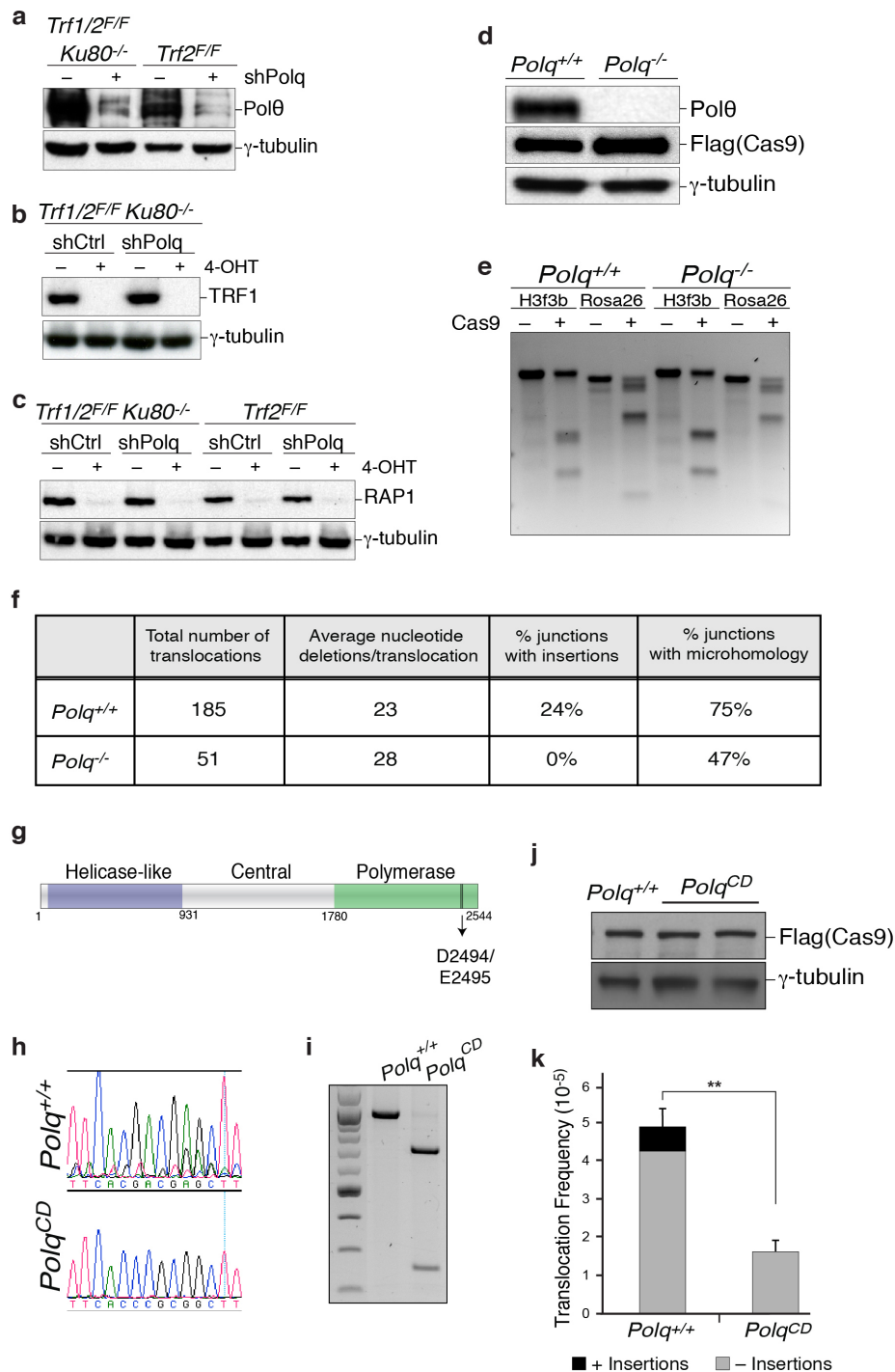
Statistical analysis. Results are presented as mean \pm s.d. of two or three independent experiments unless otherwise stated. $P < 0.05$ was considered statistically significant, and P values were calculated using the two-tailed Student's t -test. No statistical methods were used to predetermine sample size.

- Shee, C. *et al.* Engineered proteins detect spontaneous DNA breakage in human and bacterial cells. *Life* **2**, e01222 (2013).
- Leung, J. W. *et al.* Nucleosome acidic patch promotes RNF168- and RING1B/ BMI1-dependent H2AX and H2A ubiquitination and DNA damage signaling. *PLoS Genet.* **10**, e1004178 (2014).
- Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663–676 (2006).
- Prasad, R. *et al.* Human DNA polymerase θ possesses 5'-dRP lyase activity and functions in single-nucleotide base excision repair in vitro. *Nucleic Acids Res.* **37**, 1868–1877 (2009).
- Celli, G. B. & de Lange, T. DNA processing is not required for ATM-mediated telomere damage response after TRF2 deletion. *Nature Cell Biol.* **7**, 712–718 (2005).
- Sfeir, A., Kabir, S., van Overbeek, M., Celli, G. B. & de Lange, T. Loss of Rap1 induces telomere recombination in the absence of NHEJ or a DNA damage signal. *Science* **327**, 1657–1661 (2010).



Extended Data Figure 1 | Pol0 promotes alt-NHEJ repair at dysfunctional telomeres. (Related to Fig. 1.) **a**, Immunoblots for TRF1 and RAP1 after 4-OHT-induced depletion of TRF2 from *Trf2^{F/F} Cre-ERT2* MEFs and co-depletion of TRF1 and TRF2 from *Trf1^{F/F}Trf2^{F/F} Ku80^{-/-} Cre-ERT2* cells. Loss of TRF2 is confirmed by the disappearance of RAP1; a TRF2-interacting

protein the stability of which depends on TRF2 (refs 35, 36). **b**, To validate the effect of *Polq* depletion on alt-NHEJ we monitored the frequency of telomere fusions in shelterin-free *Ku80*-null cells treated with three independent shPolq vectors. shPolq-1 was used in Fig. 2. Mean values are presented with error bars denoting \pm s.e.m. from two independent experiments.

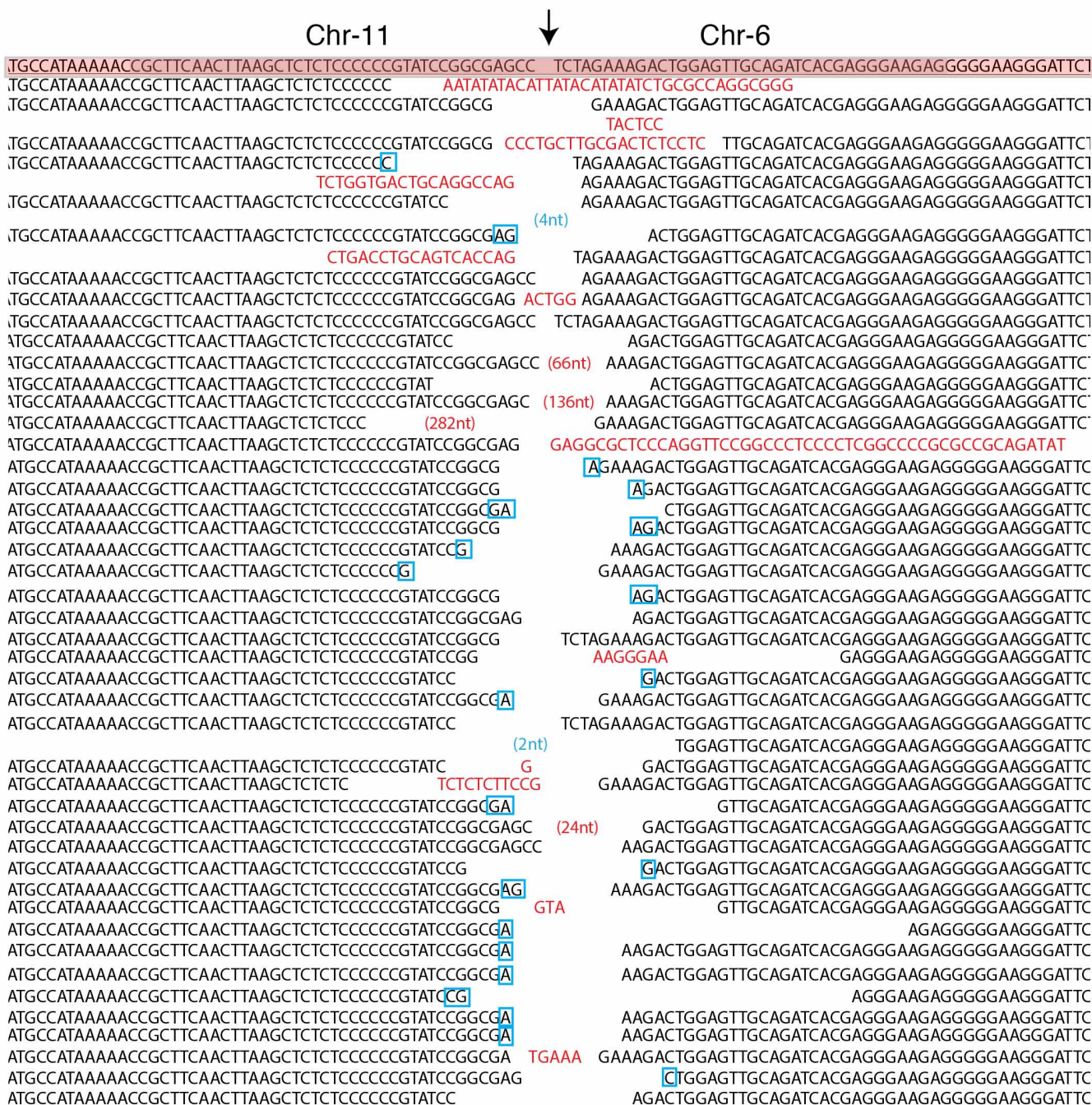


Extended Data Figure 2 | Polθ drives chromosomal translocations in mouse cells. (Related to Fig. 2.) **a**, Immunoblotting for Polθ in MEFs with the indicated genotype and treatment. **b**, Immunoblot for TRF1 in MEFs with the indicated genotype. Cells were analysed 96 h after Cre induction. **c**, RAP1 immunoblot (similar to **b**). **d**, Western blot analysis for Polθ and Flag-Cas9 in lysates prepared from *Polq*^{-/-} and *Polq*^{+/+} cells after Cas9 expression. Tubulin serves as a loading control. **e**, Surveyor nuclease assay for *Polq*^{-/-} and *Polq*^{+/+} cells expressing Cas9-gRNA(Rosa26;H3f3b) plasmid. Genomic DNA isolated from cells with the indicated genotype was used as a template to amplify across the cleavage site at either the Rosa26 or the H3f3b locus to assess intra-chromosomal NHEJ. Amplification products were denatured and then re-annealed to form heteroduplexes between unmodified and modified sequences from imprecise NHEJ. The mismatched duplex was selectively cleaved by the Surveyor nuclease at the loops that form at mismatches. **f**, Signature of translocations in *Polq*^{-/-} and *Polq*^{+/+} cells (see Extended Data Figs 3–5 for complete list of sequences). Table records the total number of

translocation events identified following CRISPR-Cas9 induced-cleavage. On average, the same number of nucleotides was deleted at the fusion junction in *Polq*^{-/-} and *Polq*^{+/+} cells. No nucleotide insertions were found in the absence of *Polq*. Lastly, the percentage of junctions exhibiting microhomology was significantly reduced in cells lacking *Polq*. **g**, Scheme depicting Polθ domains. CRISPR/Cas9 gene targeting was used to create two base substitutions at Asp2494Gly and Glu2495Ser, and generate a catalytic-dead polymerase³⁴. **h**, Sequence analysis of targeted cells. **i**, Genotyping PCRs of *Polq*^{+/+} and *Polq*^{CD} (catalytically dead allele of *Polq*) after SacI digestion. **j**, Immunoblotting to analyse Cas9 expression in *Polq*^{+/+} and two independently derived *Polq*^{CD} clonal cell lines. **k**, Frequency of chromosomal translocations (der(6)) in *Polq*^{+/+} and *Polq*^{CD} cells. Bars represent mean of four independent experiments ± s.d. (two experiments per clonal cell line). ***P* = 0.006; two-tailed Student's *t*-test. PCR products were sequenced to confirm translocation and identify possible insertions.

Translocation Junction Sequence of *Polg*^{+/+} cells

Der(11)

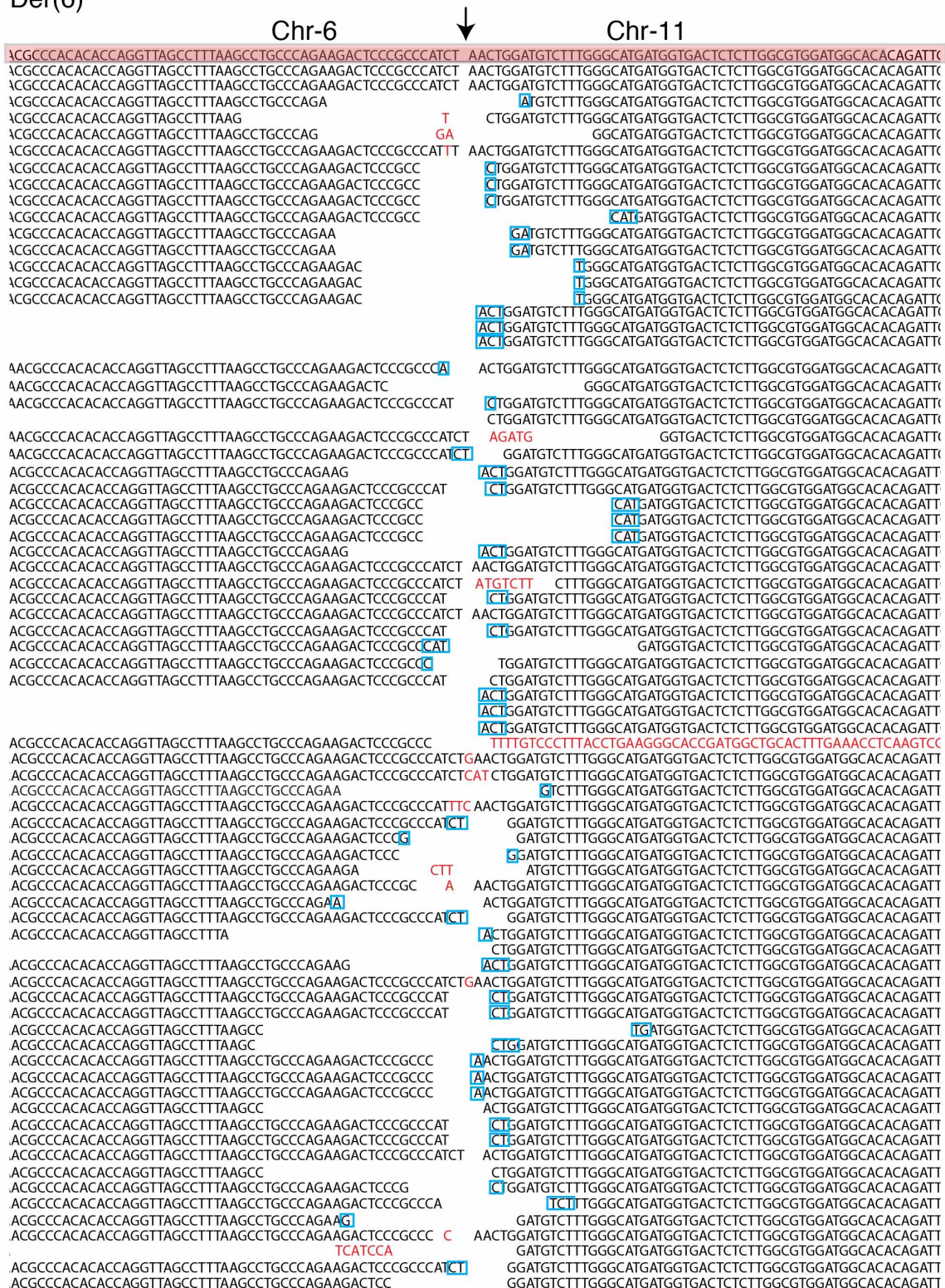


Extended Data Figure 3 | Sequence analysis of translocation junctions in *Polq*^{+/+} cells. (Related to Fig. 2.) Sequences of der(11) breakpoint junction from *Polq*^{+/+} cells. Predicted fusion breakpoint based on CRISPR cutting indicated by an arrow. Reference sequence highlighted at the top. The remaining lines represent individual translocations recovered by PCR and subject to Sanger sequencing. Nucleotide insertions are marked in red. In cases

where insertions extended beyond the sequence included in the lane, the length of the insertion was noted in parenthesis (red). Gaps in the sequence represent nucleotide deletions. The average length of the deletions was noted in Extended Data Fig. 2f. Micro-homology is denoted by blue boxes. Micro-homology embedded in DNA extending beyond the enclosed sequence was noted in parentheses (blue).

Translocation Junction Sequence of *Polq*^{+/+} cells

Der(6)

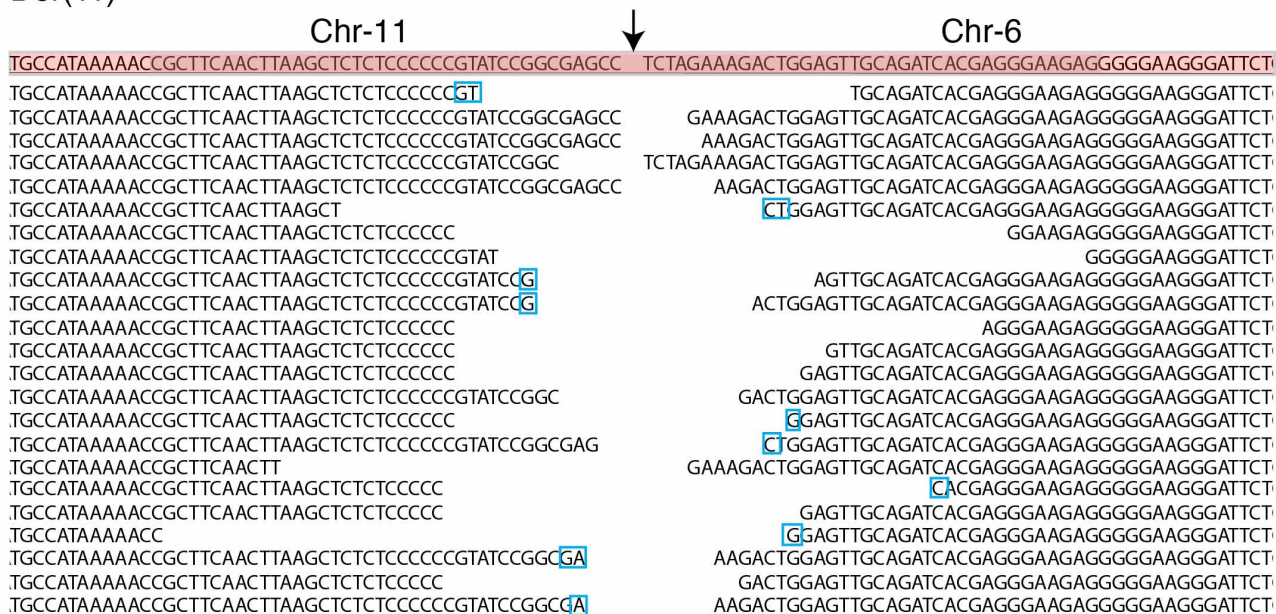


Extended Data Figure 4 | Sequence analysis of translocation junctions in *Polg*^{+/+} cells. (Related to Fig. 2.) Sequences of der(6) breakpoint junction from *Polg*^{+/+} cells. Predicted fusion breakpoint based on CRISPR cutting indicated

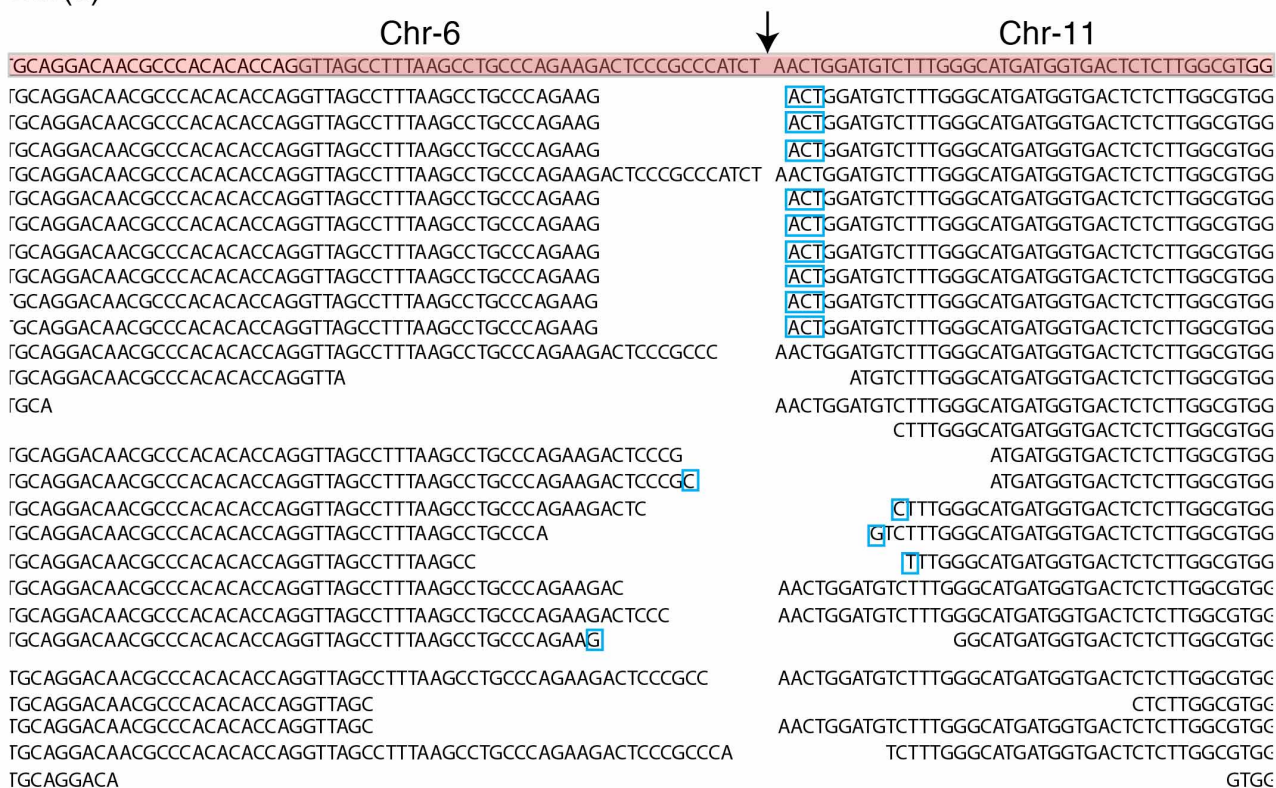
by an arrow. Reference sequence highlighted at the top. The remaining lines represent individual translocations recovered by PCR and subject to Sanger sequencing.

Translocation Junction Sequence of *Polg*^{-/-} cells

Der(11)

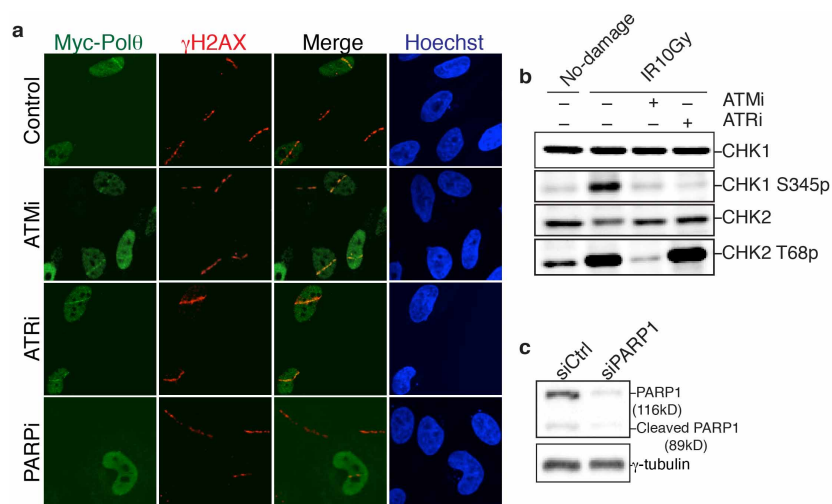


Der(6)



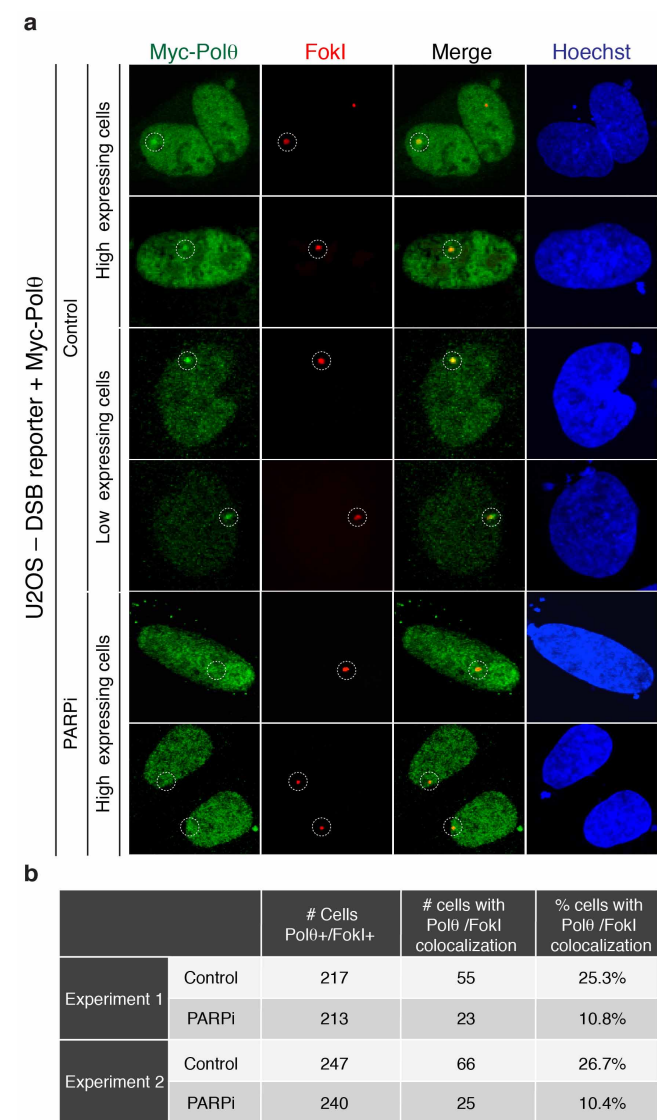
Extended Data Figure 5 | Sequence analysis of translocation junctions in *Polq*^{-/-} cells. (Related to Fig. 2.) Sequences of der(11) and der(6) breakpoint junction from *Polq*^{-/-} cells. Predicted fusion breakpoint based on CRISPR cutting indicated by an arrow. Reference sequence is highlighted at the top. The

remaining lines represent individual translocations recovered by PCR and subject to Sanger sequencing. It is important to note that insertions were completely lacking at the fusions junctions in *Polq*^{-/-} cells.

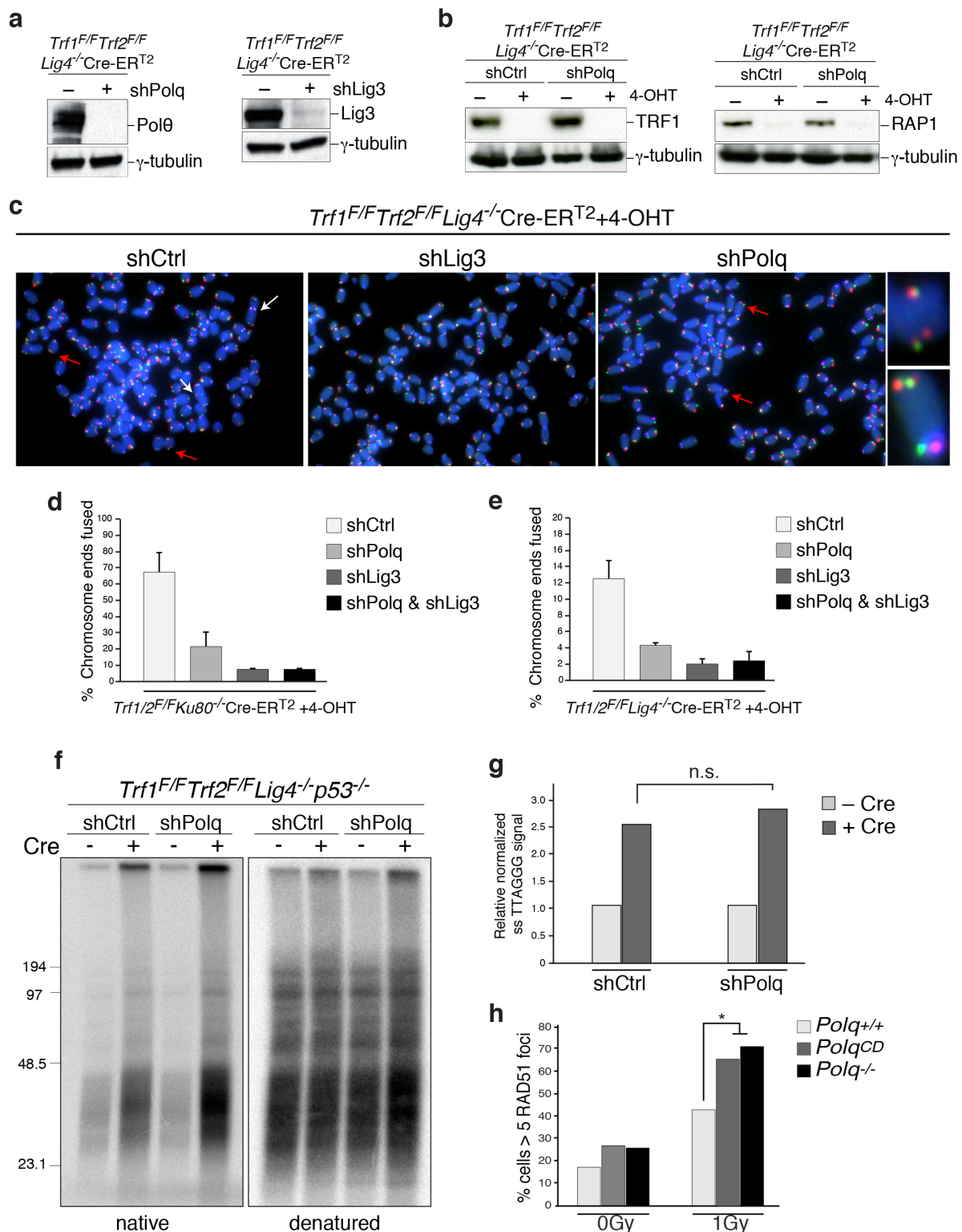


Extended Data Figure 6 | Polθ recruitment to DNA breaks. (Related to Fig. 3.) **a**, Laser micro-irradiation experiment using HeLa cells expressing Myc-Polθ and treated with ATM inhibitor (KU55933), ATR inhibitor (VE-821) or PARP inhibitor (KU58948). **b**, Western blot analysis for CHK1

and CHK2 phosphorylation. Cells with the indicated treatment were analysed 2 h after irradiation. **c**, Immunoblot for PARP1. HeLa cells were treated with PARP1 siRNA and analysed 72 h after siRNA transfection for efficiency of knockdown.

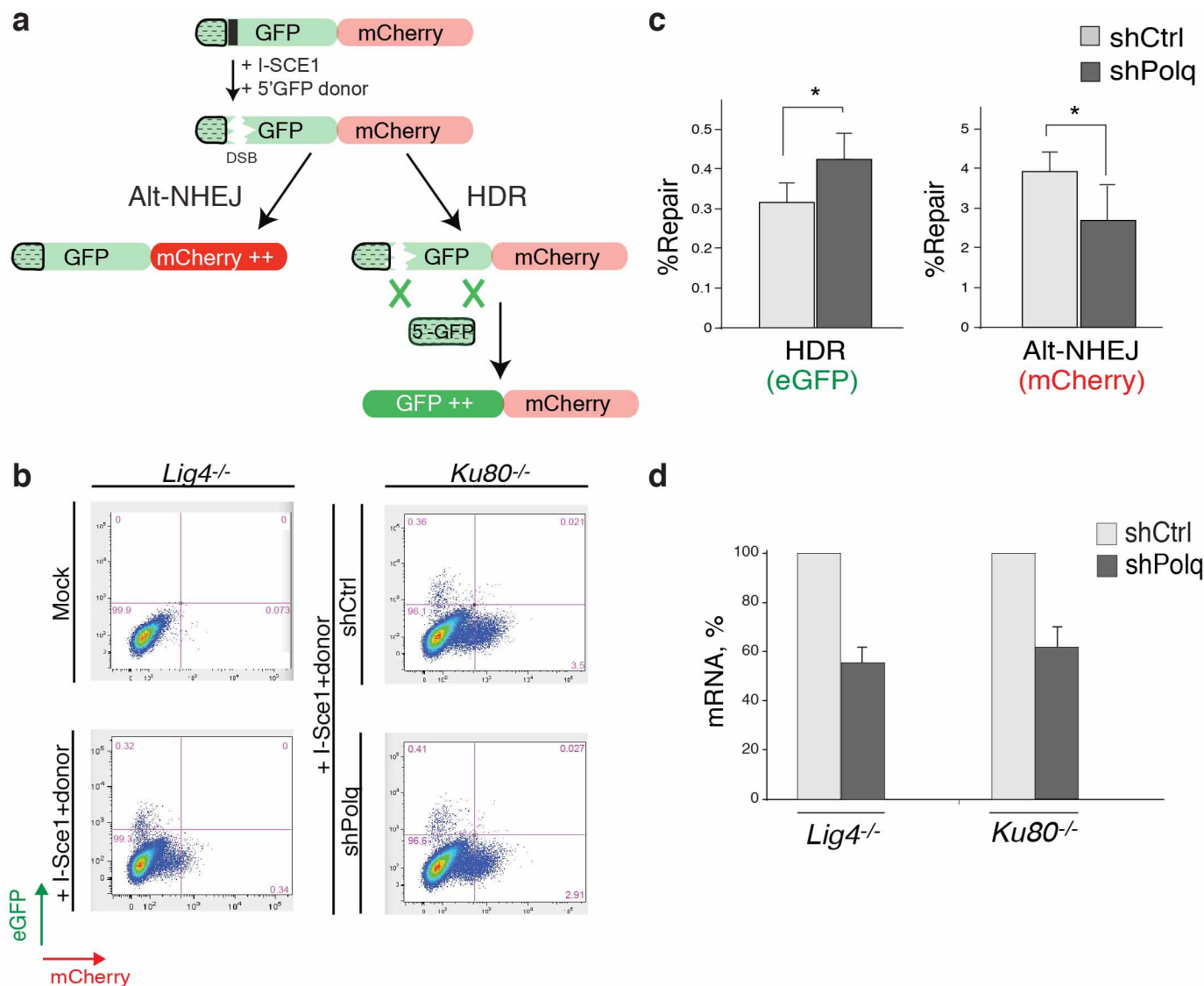


Extended Data Figure 7 | PARP1-dependent Polθ recruitment to DNA double-stranded breaks (DSBs). (Related to Fig. 3.) **a**, Results from immunofluorescence performed 4 h after induction (1 μM Shield1 ligand, Clontech 631037; 0.5 μM 4-OH tamoxifen) of DSBs by mCherry-LacI-FokI in the U2OS-DSB reporter cells¹⁸ transfected with the Myc-Polθ and treated with PARP inhibitor (KU58948). The mCherry signal is used to identify the area of damage and to assess the recruitment of Myc-Polθ to cleaved LacO repeats. **b**, Table displaying quantification related to **a**.



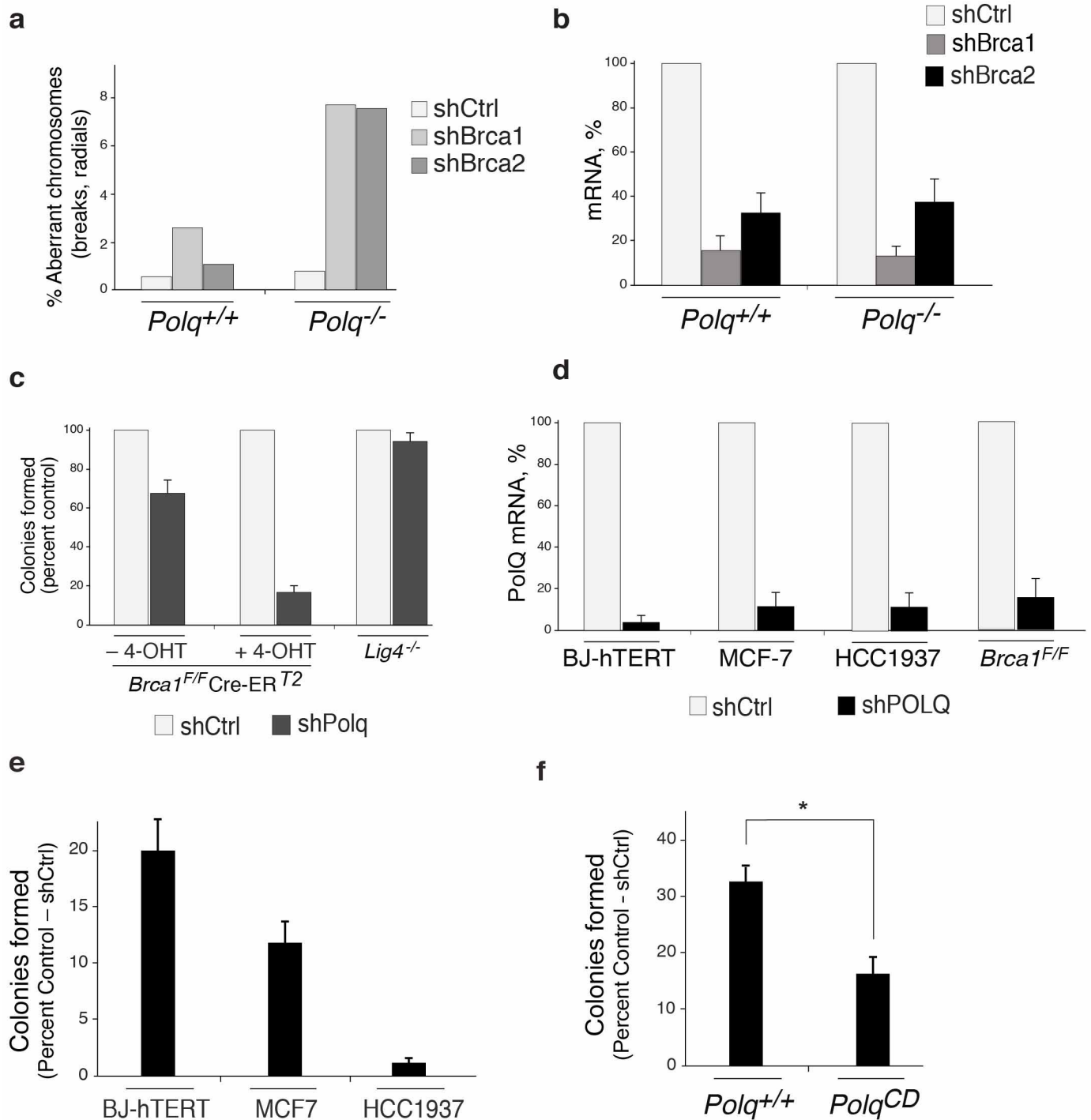
Extended Data Figure 8 | Polθ suppresses homology-directed repair at dysfunctional telomeres. (Related to Fig. 3.) **a**, Western blot analysis for Polθ and LIG3 in shelterin-free *Lig4*-null MEFs. **b**, Western blot for TRF1 and RAP1 after 4-OHT treatment of shelterin-free *Lig4*-deficient cells. **c**, Metaphase spreads from *Trf1^{F/F}Trf2^{F/F}Lig4^{-/-}Cre-ERT²* MEFs, with the indicated shRNA treatment, 96 h after Cre expression. CO-FISH assay was performed using a FITC-OO-(CCCTAA)₃ PNA probe (green) and a Tamra-OO-(TTAGGG)₃ PNA probe (red). DAPI in blue. Examples of alt-NHEJ-mediated fusion and T-SCE events (HDR) are indicated by white and red arrows, respectively. Examples of T-SCE events reflective of increased HDR in cells treated with shPolq are on the right. **d**, **e**, Quantification of telomere fusions by

alt-NHEJ in MEFs with the indicated genotype and shRNA treatment. Bars represent mean of two independent experiments \pm s.e.m. **f**, Representative in-gel hybridization to assess 3' overhang of *Trf1^{F/F}Trf2^{F/F}Lig4^{-/-}Cre-ERT²* MEFs with the indicated shRNA treatment after Cre deletion. **g**, Quantification of the gel in **f**. The single-stranded DNA/total signal ratios of the '+Cre' samples are expressed relative to the '-Cre' samples for each shRNA treatment. Mean of two independent experiments. **h**, Graph representing RAD51 accumulation after ionizing radiation treatment of *Polq^{CD}*, *Polq^{+/+}* and *Polq^{-/-}* embryonic stem cells. Bars represent mean of two independent experiments. * $P > 0.05$; two-tailed Student's *t*-test.



Extended Data Figure 9 | Polθ promotes alt-NHEJ and inhibits homology-directed repair at I-SceI-induced DNA breaks. (Related to Fig. 3.) **a**, Polθ represses recombination at DSBs induced by I-SceI. The TLR system was used to measure the relative ratio of end-joining (mCherry) and HDR (enhanced green fluorescent protein (eGFP)) repair of a DSB. A diagram of the TLR is represented. **b**, The TLR construct was stably integrated into *Lig4*^{-/-} and *Ku80*^{-/-} MEFs to avoid the confounding effect of C-NHEJ, and limit end-joining reactions to the alt-NHEJ pathway. Expression of mCherry and eGFP was assessed by flow cytometry 72 h after I-SceI and 5' eGFP donor

transduction in cells with the indicated shRNA construct. Percentages of cells are indicated in the plot. **c**, Quantification of alt-NHEJ and HDR of TLR containing *Ku80*^{-/-} MEFs after expression of I-SceI and 5' eGFP together with the indicated shRNA construct. Bar graphs represent the mean of three independent experiments \pm s.d. * $P = 0.03$; two-tailed Student's *t*-test. **d**, Real-time PCR to monitor the knockdown efficiency of *Polq* in *Ku80*^{-/-} and *Lig4*^{-/-} MEFs. The FACS analysis reported in **e** and **f** was carried out without selecting for cells expressing the shRNA-containing plasmid.



Extended Data Figure 10 | PolQ is required for survival of recombination-deficient cells. (Related to Fig. 4.) **a**, Accumulation of chromosomal aberrancies after *Brca1* and *Brca2* knockdown in *Polq*^{-/-} and *Polq*^{+/+} MEFs. Quantification of chromosomal aberrancies (chromatid breaks, chromosome breaks and radials) in MEFs stably transduced with lentiviral vectors expressing the indicated shRNA. **b**, Real-time PCR to confirm the knockdown of *Brca1* and *Brca2* as in **a**. **c**, Quantitative analysis of colony formation in *Brca1*^{F/F} Cre-ER^{T2} and *Lig4*^{-/-} cells after *Polq* depletion. The number of colonies in control shRNA-treated cells was set to 100%. Mean values are presented with error bars denoting \pm s.d. from three independent experiments. **d**, Real-time PCR to measure the knockdown efficiency of human

POLQ in BJ-hTERT, MCF7 and HCC1937 cells and mouse *Polq* in *Brca1*^{F/F} Cre-ER^{T2} cells. **e**, Quantitative analyses of colony formation in BJ-hTERT, MCF7 and HCC1937 cells after LIG3 inhibition. The number of colonies in control-shRNA-treated cells was set to 100%. The knockdown efficiency for *Lig3* was \sim 85%. Bars represent mean of two independent experiments \pm s.e.m. **f**, Quantitative analyses of colony formation in *Polq*^{CD} and *Polq*^{+/+} embryonic stem cells after *BRCA1* inhibition. The number of colonies in control-shRNA-treated cells was set to 100%. The knockdown efficiency for *BRCA1* was $>$ 80%. Bars represent mean of two independent experiments \pm s.e.m. $*P = 0.05$; two-tailed Student's *t*-test.

Homologous-recombination-deficient tumours are dependent on Polθ-mediated repair

Raphael Ceccaldi¹, Jessica C. Liu^{1,2,3}, Ravindra Amunugama⁴, Ildiko Hajdu⁵, Benjamin Primack¹, Mark I. R. Petalcorin⁶, Kevin W. O'Connor¹, Panagiotis A. Konstantinopoulos⁷, Stephen J. Elledge⁵, Simon J. Boulton⁶, Timur Yusufzai^{1,2} & Alan D. D'Andrea¹

Large-scale genomic studies have shown that half of epithelial ovarian cancers (EOCs) have alterations in genes regulating homologous recombination (HR) repair¹. Loss of HR accounts for the genomic instability of EOCs and for their cellular hyper-dependence on alternative poly-ADP ribose polymerase (PARP)-mediated DNA repair mechanisms^{2–5}. Previous studies have implicated the DNA polymerase θ (Polθ also known as POLQ, encoded by *POLQ*)⁶ in a pathway required for the repair of DNA double-strand breaks^{7–9}, referred to as the error-prone microhomology-mediated end-joining (MMEJ) pathway^{10–13}. Whether Polθ interacts with canonical DNA repair pathways to prevent genomic instability remains unknown. Here we report an inverse correlation between HR activity and Polθ expression in EOCs. Knockdown of Polθ in HR-proficient cells upregulates HR activity and RAD51 nucleofilament assembly, while knockdown of Polθ in HR-deficient EOCs enhances cell death. Consistent with these results, genetic inactivation of an HR gene (*Fancd2*) and *Polq* in mice results in embryonic lethality. Moreover, Polθ contains RAD51 binding motifs and it blocks RAD51-mediated recombination. Our results reveal a synthetic lethal relationship between the HR pathway and Polθ-mediated repair in EOCs, and identify Polθ as a novel druggable target for cancer therapy.

To examine changes in polymerase activity between tumours and normal tissues, we screened polymerase gene expression profiles in a large number of cancers (Supplementary Table 1). Gene set enrichment analysis (GSEA) revealed specific and recurrent overexpression of Polθ in EOCs (Extended Data Fig. 1a–c). Polθ was upregulated in a grade-dependent manner and its expression positively correlated with numerous mediators of HR (Extended Data Fig. 1d–j). As Polθ has been suggested to play a role in DNA repair^{7–10}, we investigated a potential role for Polθ in HR repair.

To test the relationship between Polθ expression and HR, we used a cell-based assay in human cells which measures the efficiency of recombination of two GFP alleles (DR-GFP assay)¹⁴. Knockdown of Polθ with short interfering RNA (siRNA) (Extended Data Fig. 2a) resulted in an increase in HR efficiency, similar to that observed by depleting the anti-recombinases PARI or BLM^{15,16}. Depletion of Polθ caused a significant increase in basal and radiation (IR)-induced RAD51 foci (Fig. 1a, b and Extended Data Fig. 2b–d), and depletion of Polθ in 293T cells conferred cellular hypersensitivity to mitomycin C (MMC) and an increase in MMC-induced chromosomal aberrations (Extended Data Fig. 2e, f). These findings suggest that human Polθ inhibits HR and participates in the maintenance of genome stability.

Given that Polθ shares structural homology with coexpressed RAD51-binding ATPases (Extended Data Fig. 1k, l), we hypothesized that Polθ might regulate HR through an interaction with RAD51. RAD51 was detected in Flag-tagged Polθ immunoprecipitates, and purified full-length

Flag-Polθ bound recombinant human RAD51 (Fig. 1c, d). Pull-down assays with recombinant GST-RAD51 and *in vitro* translated Polθ truncation mutants defined a region of Polθ binding to RAD51 spanning amino acids 847–894 (Fig. 1e, f and Extended Data Fig. 2g, h). Sequence homology of Polθ with the RAD51 binding domain of *C. elegans* RFS-1 (ref. 17) identified a second binding region (Extended Data Fig. 2i). Peptides arrays narrowed down the RAD51 binding activity of Polθ to three distinct motifs (Fig. 1g and Extended Data Fig. 2j). Substitution arrays confirmed the interaction and highlighted the importance of the 847–894 Polθ region as both necessary and sufficient for RAD51 binding (Extended Data Fig. 3a, b). Together these results indicate that Polθ is a RAD51-interacting protein that regulates HR.

In order to address the role of Polθ in HR regulation, we assessed the ability of wild-type or mutant Polθ to complement the siPolθ-dependent increase in RAD51 foci. Full-length wild-type Polθ fully reduced IR-induced RAD51 foci, unlike Polθ mutated at ATPase catalytic residues (A-dead) or Polθ lacking interaction with RAD51 (ΔRAD51) (Fig. 2a, b). Expression of a Polθ mutant lacking the polymerase domain (ΔPol1) was sufficient to decrease IR-induced RAD51 foci, suggesting that the N-terminal half of Polθ is sufficient to disrupt RAD51 foci (Fig. 2b and Extended Data Fig. 3c, d). We next measured the ability of wild-type or mutant Polθ to complement the siPolθ-dependent increase in HR efficiency. Again, expression of full-length Polθ or ΔPol1 decreased the recombination frequency when compared to cells expressing other Polθ constructs, suggesting that the N-terminal half of Polθ containing the RAD51 binding domain and the ATPase domain is needed to inhibit HR (Fig. 2c and Extended Data Fig. 3e).

A purified recombinant Polθ fragment (ΔPol2) from insect cells exhibited low levels of basal ATPase activity, as previously reported¹⁸ (Fig. 2d, e). Polθ ATPase activity was selectively stimulated by the addition of single-stranded DNA (ssDNA) or fork DNA (Fig. 2e and Extended Data Fig. 4a). Electrophoretic mobility gel shift assays (EMSA) showed specific binding of Polθ to ssDNA (Fig. 2f and Extended Data Fig. 4b). We incubated ΔPol2 with ssDNA and measured RAD51–ssDNA nucleofilament assembly. Interestingly, RAD51–ssDNA assembly was reduced by ΔPol2 wild-type but not by A-dead or ΔRAD51, indicating that Polθ negatively affects RAD51–ssDNA assembly through its RAD51 binding and ATPase activities (Fig. 2g and Extended Data Fig. 4c–f). Furthermore, Polθ decreased the efficiency of D-loop formation, confirming that Polθ is a negative regulator of HR (Fig. 2h and Extended Data Fig. 4g–j).

As Polθ is upregulated in subgroups of cancers associated with HR deficiency (Fig. 3a) and Polθ activity shows specificity for replicative-stress-mediated structures (ssDNA and fork DNA) (Fig. 2e, f), we examined the cellular functions of Polθ under replicative stress. Subcellular fractionation revealed that Polθ is enriched in chromatin in response to ultraviolet (UV) light; and RAD51 binding by Polθ was enhanced by UV

¹Department of Radiation Oncology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, Massachusetts 02215, USA. ²Department of Biological Chemistry & Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02215, USA. ³Department of Molecular & Cellular Biology, Harvard University, Cambridge, Massachusetts 02138, USA. ⁴Howard Hughes Medical Institute, Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02215, USA. ⁵Howard Hughes Medical Institute, Division of Genetics, Brigham and Women's Hospital, Boston, Massachusetts 02215, USA. ⁶DNA Damage Response Laboratory, Cancer Research UK, London Research Institute, Clare Hall, South Mimms EN6 3LD, UK. ⁷Department of Medical Oncology, Medical Gynecologic Oncology Program, Dana-Farber Cancer Institute, Harvard Medical School, Boston, Massachusetts 02215, USA.

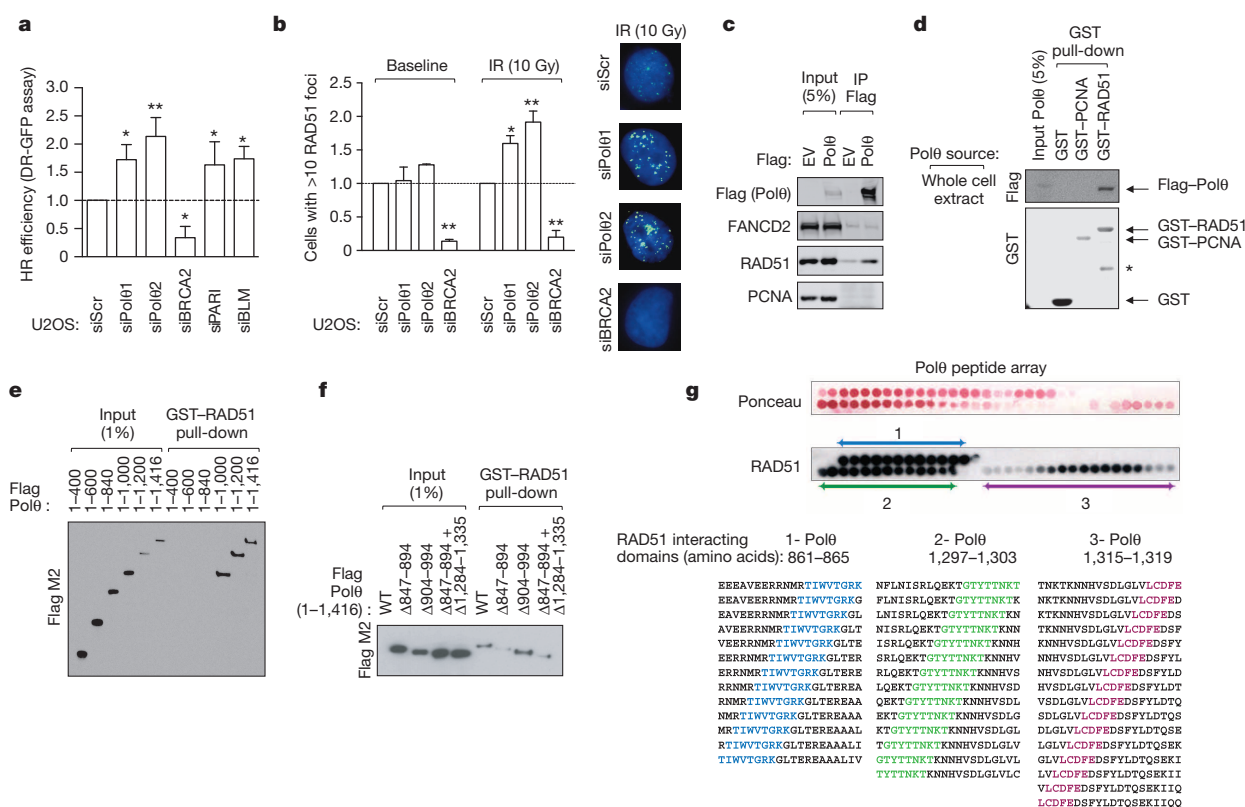


Figure 1 | Polθ is a RAD51-interacting protein that suppresses HR.

a, DR-GFP assay in U2OS cells transfected with indicated siRNA. **b**, Quantification of RAD51 foci in U2OS cells transfected with indicated siRNA. **c**, Endogenous RAD51 co-precipitates *in vivo* with purified full-length Flag-tagged Polθ from whole cell extracts. EV, empty vector. **d**, GST pull-down experiment with full-length Flag-tagged Polθ (* indicates non-specific band).

e, GST–RAD51 pull-down with *in vitro* translated Polθ truncation mutants. **f**, GST–RAD51 pull-down with *in vitro* translated Polθ versions missing indicated amino acids. **g**, Ponceau staining and immunoblotting of peptide arrays for the indicated Polθ motifs probed with recombinant RAD51. The Polθ amino acids spanning RAD51-interacting motifs are shown. Data in **a** and **b** represent mean \pm s.e.m.

exposure, suggesting that Polθ regulates HR in cells under replicative stress (Extended Data Fig. 5a, b). Polθ-depleted cells were hypersensitive to cellular stress and DNA damage, along with an exacerbated checkpoint

activation and increased γ H2AX phosphorylation (Fig. 3b, c). Furthermore, the cell cycle progression of Pol θ -depleted cells was impaired after DNA damage (Fig. 3d, e). To determine the role of Pol θ in replication

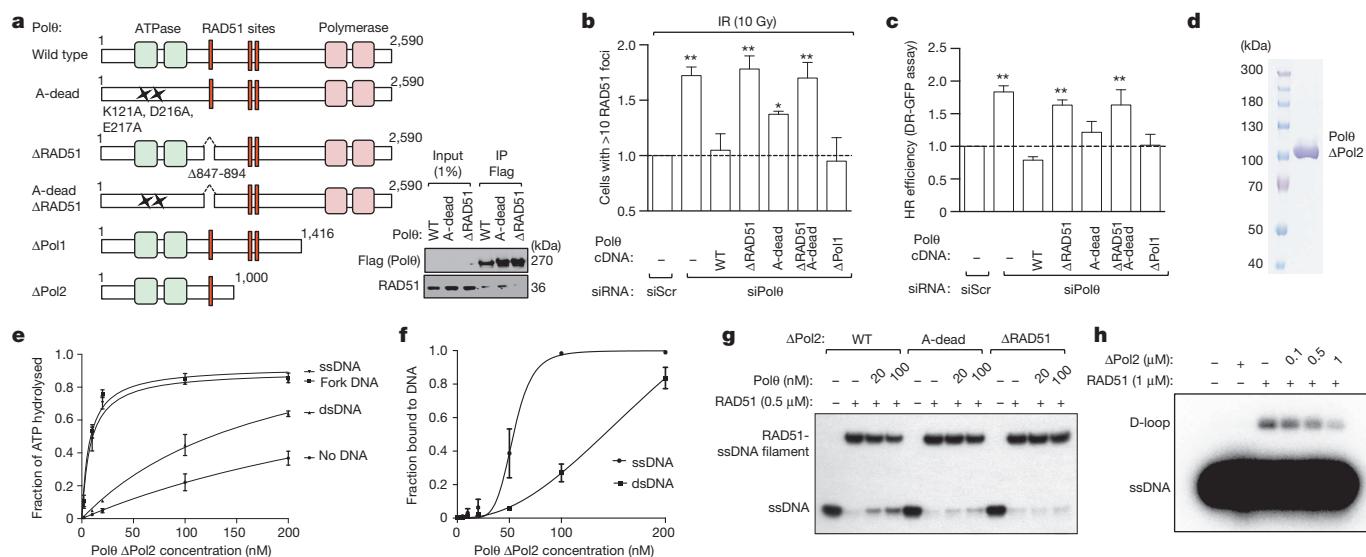


Figure 2 | Polθ inhibits RAD51-mediated recombination. **a**, Schematic of Polθ mutants used in complementation studies and their interaction with RAD51. WT, wild type. **b**, Quantification of RAD51 foci in U2OS cells transfected with indicated siRNA and Polθ cDNA constructs refractory to siPolθ1. **c**, DR-GFP assay in U2OS cells transfected with indicated siRNA and

Polθ cDNA constructs refractory to siPolθ1. **d**, Coomassie-stained gel of the purified Polθ fragment. **e**, Quantification of Polθ ATPase activity. **f**, Quantification of Polθ binding to ssDNA and dsDNA. **g**, RAD51–ssDNA nucleofilament assembly assay. **h**, Assessment of RAD51-dependent D-loop formation. Data in **b**, **c**, **e** and **f** represent mean \pm s.e.m.

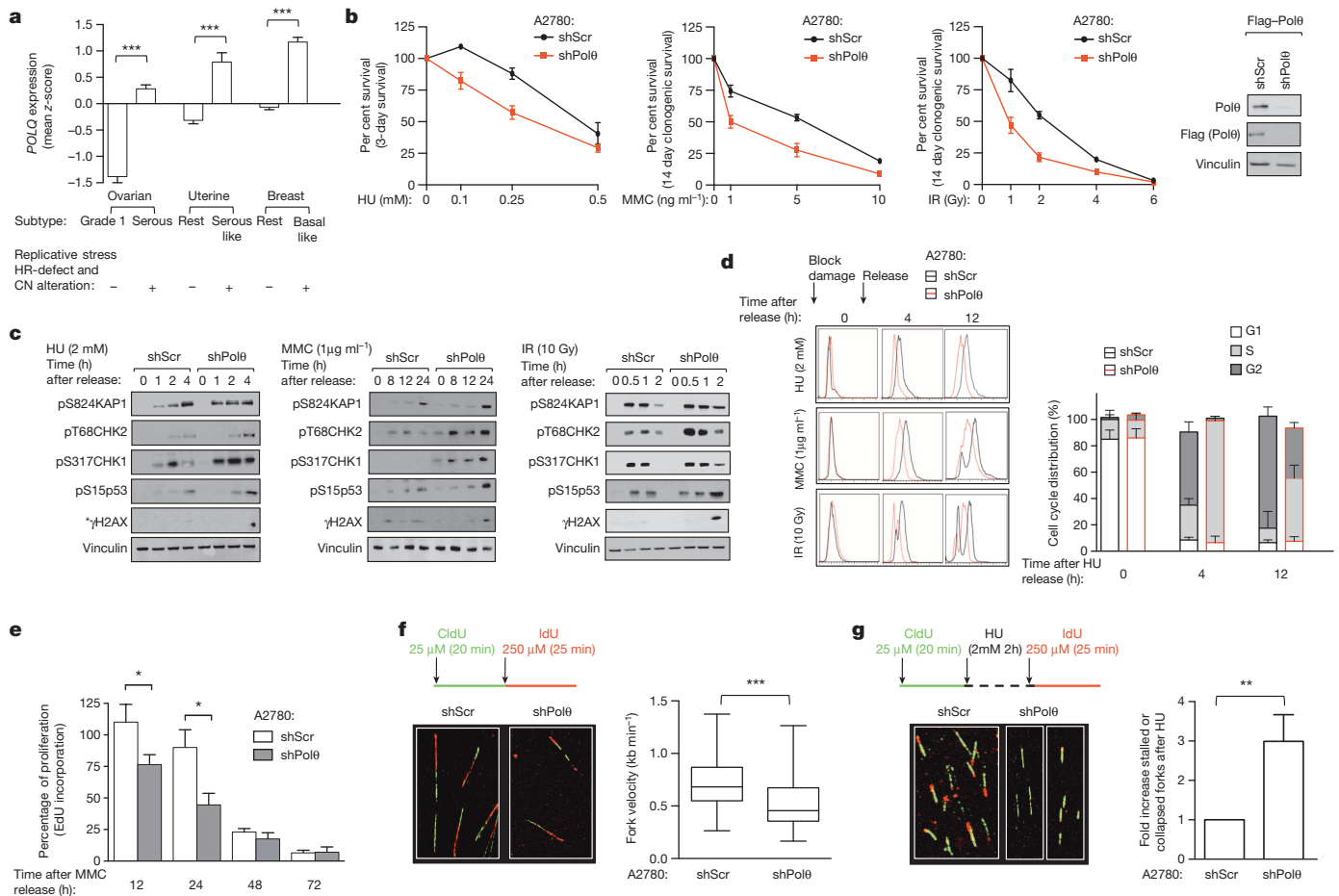


Figure 3 | Polθ promotes S phase progression and recovery of stalled forks.

a, POLQ gene expression in subtypes of cancers with HR deficiency. **b**, Survival assays of A2780 cells exposed to the indicated DNA-damaging agents. Immunoblot showing silencing efficiency is shown on the right. **c**, Immunoblot analyses following pulse treatments with DNA-damaging agents (*γH2AX: see Methods for specific time points used for γH2AX immunoblot). HU, hydroxyurea. **d**, Cell cycle progression of synchronized A2780 cells. **e**, Fraction of cycling A2780 cells measured by 5-ethynyl-2'-deoxyuridine (EdU) incorporation. **f**, Quantification of DNA fibre lengths. **g**, Percentage of stalled forks. All experiments shown in **a–d** were performed in two cell lines (A2780 and 293T). All data represent mean ± s.e.m. except for box plots in **f** that show twenty-fifth to seventy-fifth percentiles, with lines indicating the median, and whiskers indicating the smallest and largest values.

dynamics, single-molecule analyses were performed on extended DNA fibres¹⁹. Abnormalities in replication fork progression were observed in Polθ-depleted cells (Fig. 3f, g and Extended Data Fig. 5c, d). These results suggest that Polθ maintains genomic stability at stalled or collapsed replication forks by promoting fork restart.

To examine the regulation of Polθ, we quantified Polθ expression by RT-qPCR. Polθ was selectively upregulated in HR-deficient ovarian cancer cell lines. Complementation of BRCA1 or FANCD2-deficient cell lines with BRCA1 and FANCD2 cDNA respectively, restored normal HR function and reduced Polθ expression to normal levels. Conversely, siRNA-mediated inhibition of HR genes increased Polθ expression (Extended Data Fig. 5e, f). Polθ expression was significantly higher in subgroups of cancers with HR deficiency and a high genomic instability pattern²⁰ (Fig. 3a and Extended Data Fig. 5g). Patients with high Polθ expression had a better response to platinum chemotherapy, a surrogate for HR deficiency, suggesting that Polθ expression inversely correlates with HR activity and may be useful as a biomarker for platinum sensitivity (Extended Data Fig. 5h, i). Together, these data indicate that increased Polθ expression is driven by HR deficiency.

To assess the possible synthetic lethality between HR genes and Polθ, we generated an HR-deficient ovarian tumour cell line, A2780-shFANCD2 cells (Extended Data Fig. 6a–c). These cells, and the parental A2780 cells, were subjected to Polθ depletion, and survival following exposure to cytotoxic drugs was measured. Polθ depletion reduced the survival of

representative cell cycle distribution. **e**, Fraction of cycling A2780 cells measured by 5-ethynyl-2'-deoxyuridine (EdU) incorporation. **f**, Quantification of DNA fibre lengths. **g**, Percentage of stalled forks. All experiments shown in **a–d** were performed in two cell lines (A2780 and 293T). All data represent mean ± s.e.m. except for box plots in **f** that show twenty-fifth to seventy-fifth percentiles, with lines indicating the median, and whiskers indicating the smallest and largest values.

HR-deficient cells exposed to inhibitors of PARP (PARPi), cisplatin (CDDP) or MMC (Extended Data Fig. 6d–f). Polθ inhibition impaired the survival of BRCA1-deficient tumours (MDA-MB-436) after PARPi treatment but had no effect on the complemented line (MDA-MB-436 + BRCA1) (Fig. 4a). Polθ-depleted cells were hypersensitive to ATM inhibition, known to create an HR defect phenotype²¹. Chromosomal breakage, checkpoint activation, and γH2AX phosphorylation in response to MMC were exacerbated by Polθ depletion (Fig. 4b and Extended Data Fig. 6g, h). Furthermore, a whole-genome short hairpin RNA (shRNA) screen performed on HR-deficient (*FANCA*^{-/-}) fibroblasts showed that shRNAs targeting Polθ impair cell survival in MMC (Extended Data Fig. 6i), suggesting that HR-deficient cells cannot survive in the absence of Polθ.

Next, we investigated the interaction between the HR and Polθ pathways *in vivo* by interbreeding *Fancd2*^{+/-} and *Polq*^{+/-} mice. Although *Fancd2*^{-/-} and *Polq*^{-/-} mice are viable and exhibit subtle phenotypes^{7,22}, viable *Fancd2*^{-/-}*Polq*^{-/-} mice were uncommon from these matings (Extended Data Fig. 7a). The only surviving *Fancd2*^{-/-}*Polq*^{-/-} pups exhibited severe congenital malformations and were either found dead or died prematurely. *Fancd2*^{-/-}*Polq*^{-/-} embryos showed severe congenital malformations, and mouse embryonic fibroblasts (MEFs) generated from *Fancd2*^{-/-}*Polq*^{-/-} embryos showed hypersensitivity to PARPi (Fig. 4c and Extended Data Fig. 7b–e). These data suggest that loss of the HR and Polθ repair pathways *in vivo* results in embryonic lethality.

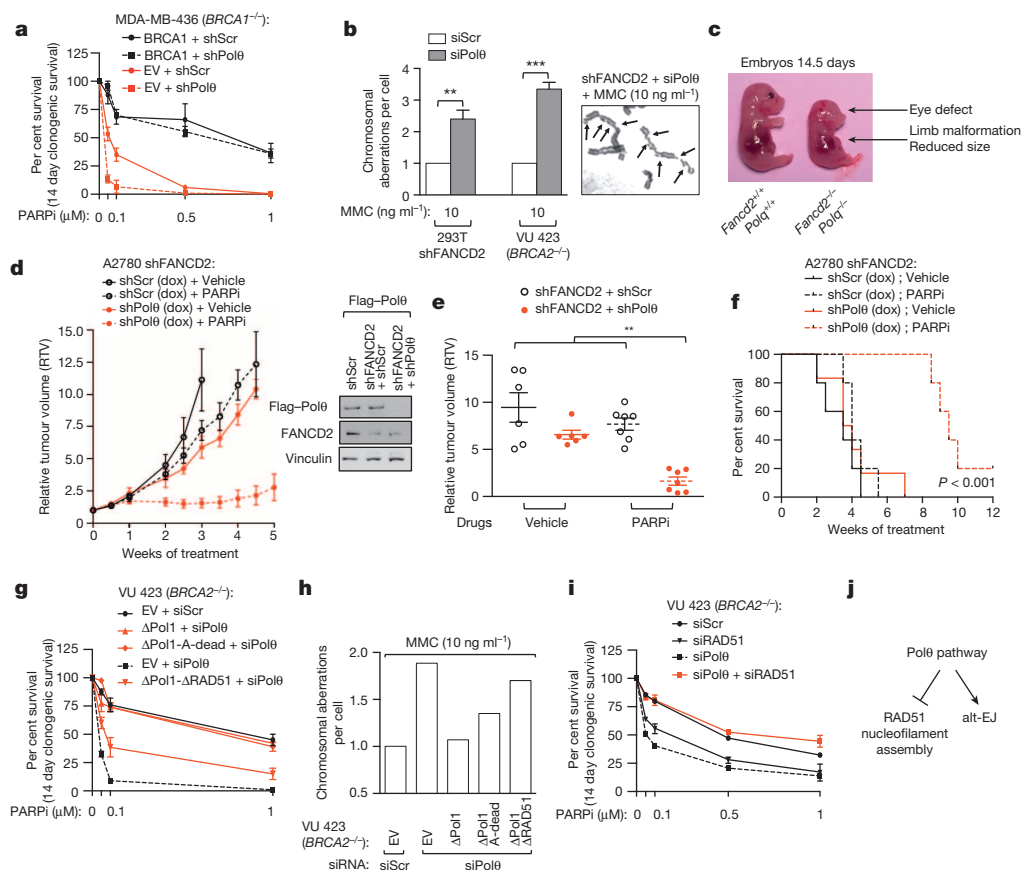


Figure 4 | Synthetic lethality between HR and Polθ repair pathways.

a, Clonogenic formation of BRCA1-deficient (MDA-MB-436) cells expressing indicated cDNA together with indicated shRNA. **b**, Chromosome breakage analysis of HR-deficient cells transfected with the indicated siRNA. A representative image is shown. Arrows indicate chromosomal aberrations. **c**, Embryos at day 14 of gestation. **d**, Growth of indicated xenografts *in vivo*. Immunoblot showing silencing efficiency. **e**, Relative tumour volumes (RTV) for individual mice treated in **d** after three weeks of treatment. **f**, Overall

survival for mice treated with vehicle or PARPi. Log-rank $P < 10^{-3}$.

g, **h**, Clonogenic formation (**g**) and chromosome breakage analysis (**h**) of BRCA2-deficient cells expressing Polθ cDNA constructs refractory to siPolθ1 and transfected with the indicated siRNA. **i**, Clonogenic formation of BRCA2-deficient cells transfected with the indicated siRNA. **j**, Model for role of Polθ in DNA repair. Data in **a**, **b**, **g** and **i** represent mean \pm s.e.m. For data in **d**–**f**, each circle represents data from one tumour and each group represents $n \geq 7$ tumours from $n \geq 6$ mice. Brackets show mean \pm s.e.m.

As xenografts of tumours cells expressing shRNAs against both FANCD2 and Polθ did not stably propagate in mice (Extended Data Fig. 7f), we xenotransplanted A2780-shFANCD2 cells expressing either doxycycline-inducible Polθ or scrambled (Scr) shRNA in athymic nude mice. Polθ depletion significantly impaired tumour growth after PARPi treatment (Fig. 4d, e and Extended Data Fig. 7g, h). Moreover, mice bearing Polθ-depleted tumours had a survival advantage following PARPi treatment compared to control mice (Fig. 4f). Polθ-depleted HR-deficient tumour cells also exhibited decreased survival in *in vivo* dual-colour competition experiments (Extended Data Fig. 7i–l). Collectively, these data confirm that HR-deficient tumours are hypersensitive to inhibition of Polθ-mediated repair.

To understand which functions of Polθ are required for resistance to DNA-damaging agents, we performed a series of complementation studies in HR-deficient cells. Expression of full-length Polθ or ΔPol1, but not ΔRAD51, in HR-deficient Polθ-depleted cells treated with PARPi or MMC was able to rescue toxicity, suggesting that the anti-recombinase activity of Polθ maintains the genomic stability of HR-deficient cells (Fig. 4g, h and Extended Data Fig. 8a, b). Moreover, the toxicity induced by loss of Polθ in HR-deficient cells was rescued by depletion of RAD51 showing that, in the absence of Polθ, RAD51 is toxic to HR-deficient cells (Fig. 4i). These results suggest a role for Polθ in limiting toxic HR events²³ (Extended Data Fig. 8c–f) and may explain why HR-deficient cells overexpress and depend on an anti-recombinase for survival.

High mutation rates have been observed in HR-deficient tumours²⁴. Previous studies have shown that Polθ is an error-prone polymerase^{25,26}

that participates in alternative end-joining (alt-EJ)¹⁰. Therefore, we assessed the role of Polθ in error-prone DNA repair in human cancer cells. Polθ inhibition reduced alt-EJ efficiency in U2OS cells, similar to the reduction observed following depletion of PARP1, another critical factor in end-joining^{27,28} (Extended Data Fig. 9a). Expression of full-length Polθ, ΔRAD51, or A-dead, but not the ΔPol1 mutant, complemented the cells, suggesting that the polymerase domain of Polθ is required for end-joining (Extended Data Fig. 9b). GFP-tagged full-length Polθ formed foci after UV treatment in a PARP1-dependent manner (Extended Data Fig. 9c). Polθ inhibition reduced the mutation frequency induced by UV light, and tumours with high Polθ expression harboured more somatic point mutations than those with lower Polθ levels (Extended Data Fig. 9d, e). These results suggest that Polθ contributes to the mutational signature observed in some HR-deficient tumours²⁹.

In human cancers, a deficiency in one DNA repair pathway can result in cellular hyper-dependence on a second compensatory DNA repair pathway⁴. Here we show that Polθ is overexpressed in EOCs and other tumours with HR defects³⁰. Wild-type Polθ limits RAD51–ssDNA nucleofilament assembly (Extended Data Fig. 10a) and promotes alt-EJ (Fig. 4j). We demonstrate that HR-deficient tumours are hypersensitive to inhibition of Polθ-mediated repair. Therefore, Polθ appears to channel DNA repair by antagonizing HR and promoting PARP1-dependent error-prone repair (Extended Data Fig. 10b). These results offer a potential new therapeutic target for cancers with inactivated HR.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 1 September; accepted 24 December 2014.

Published online 2 February 2015.

1. The Cancer Genome Atlas Research Network. Integrated genomic analyses of ovarian carcinoma. *Nature* **474**, 609–615 (2011).
2. Farmer, H. *et al.* Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. *Nature* **434**, 917–921 (2005).
3. Bryant, H. E. *et al.* Specific killing of BRCA2-deficient tumours with inhibitors of poly(ADP-ribose) polymerase. *Nature* **434**, 913–917 (2005).
4. Kennedy, R. D. & D'Andrea, A. D. DNA repair pathways in clinical practice: lessons from pediatric cancer susceptibility syndromes. *J. Clin. Oncol.* **24**, 3799–3808 (2006).
5. Bast, R. C. Jr, Hennessey, B. & Mills, G. B. The biology of ovarian cancer: new opportunities for translation. *Nature Rev. Cancer* **9**, 415–428 (2009).
6. Yousefzadeh, M. J. & Wood, R. D. DNA polymerase POLQ and cellular defense against DNA damage. *DNA Repair* **12**, 1–9 (2013).
7. Shima, N., Munroe, R. J. & Schimenti, J. C. The mouse genomic instability mutation *chaos1* is an allele of *Polq* that exhibits genetic interaction with *Atm*. *Mol. Cell. Biol.* **24**, 10381–10389 (2004).
8. Yoshimura, M. *et al.* Vertebrate POLQ and POL β cooperate in base excision repair of oxidative DNA damage. *Mol. Cell* **24**, 115–125 (2006).
9. Muzzini, D. M., Plevani, P., Boulton, S. J., Cassata, G. & Marini, F. *Caenorhabditis elegans* POLQ-1 and HEL-308 function in two distinct DNA interstrand cross-link repair pathways. *DNA Repair* **7**, 941–950 (2008).
10. McVey, M. & Lee, S. E. MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings. *Trends in Genet.* **24**, 529–538 (2008).
11. Chan, S. H., Yu, A. M. & McVey, M. Dual roles for DNA polymerase theta in alternative end-joining repair of double-strand breaks in *Drosophila*. *PLoS Genet.* **6**, e1001005 (2010).
12. Yu, A. M. & McVey, M. Synthesis-dependent microhomology-mediated end joining accounts for multiple types of repair junctions. *Nucleic Acids Res.* **38**, 5706–5717 (2010).
13. Koole, W. *et al.* A polymerase theta-dependent repair pathway suppresses extensive genomic instability at endogenous G4 DNA sites. *Nat. Commun.* **5**, 3216 (2014).
14. Nakanishi, K. *et al.* Human Fanconi anemia monoubiquitination pathway promotes homologous DNA repair. *Proc. Natl Acad. Sci. USA* **102**, 1110–1115 (2005).
15. Moldovan, G. L. *et al.* Inhibition of homologous recombination by the PCNA-interacting protein PARI. *Mol. Cell* **45**, 75–86 (2012).
16. Ira, G., Malkova, A., Liberi, G., Foiani, M. & Haber, J. E. Srs2 and Sgs1-Top3 suppress crossovers during double-strand break repair in yeast. *Cell* **115**, 401–411 (2003).
17. Ward, J. D. *et al.* Overlapping mechanisms promote postsynaptic RAD-51 filament disassembly during meiotic double-strand break repair. *Mol. Cell* **37**, 259–272 (2010).
18. Seki, M., Marini, F. & Wood, R. D. POLQ (Pol theta), a DNA polymerase and DNA-dependent ATPase in human cells. *Nucleic Acids Res.* **31**, 6117–6126 (2003).
19. Jackson, D. A. & Pombo, A. Replicon clusters are stable units of chromosome structure: evidence that nuclear organization contributes to the efficient activation and propagation of S phase in human cells. *J. Cell Biol.* **140**, 1285–1295 (1998).
20. The Cancer Genome Atlas Research Network. Integrated genomic characterization of endometrial carcinoma. *Nature* **497**, 67–73 (2013).
21. Morrison, C. *et al.* The controlling role of ATM in homologous recombinational repair of DNA damage. *EMBO J.* **19**, 463–471 (2000).
22. Parmar, K. *et al.* Hematopoietic stem cell defects in mice with deficiency of Fancd2 or Usp1. *Stem Cells* **28**, 1186–1195 (2010).
23. Hu, Y. *et al.* PARP1-driven poly-ADP-ribosylation regulates BRCA1 function in homologous recombination mediated DNA repair. *Cancer Discov.* **4**, 1430–1447 (2014).
24. Yang, D. *et al.* Association of BRCA1 and BRCA2 mutations with survival, chemotherapy sensitivity, and gene mutator phenotype in patients with ovarian cancer. *J. Am. Med. Assoc.* **306**, 1557–1565 (2011).
25. Seki, M. *et al.* High-efficiency bypass of DNA damage by human DNA polymerase θ . *EMBO J.* **23**, 4484–4494 (2004).
26. Seki, M. & Wood, R. D. DNA polymerase theta (POLQ) can extend from mismatches and from bases opposite a (6–4) photoproduct. *DNA Repair* **7**, 119–127 (2008).
27. Zhang, Y. & Jasin, M. An essential role for CtIP in chromosomal translocation formation through an alternative end-joining pathway. *Nature Struct. Mol. Biol.* **18**, 80–84 (2011).
28. Chiruvella, K. K., Liang, Z. & Wilson, T. E. Repair of double-strand breaks by end joining. *Cold Spring Harb. Perspect. Biol.* **5**, a012757 (2013).
29. Alexandrov, L. B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 (2013).
30. Lemée, F. *et al.* DNA polymerase theta up-regulation is associated with poor survival in breast cancer, perturbs DNA replication, and promotes genetic instability. *Proc. Natl Acad. Sci. USA* **107**, 13390–13395 (2010).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank N. Shima, D. Chowdhury, G. Shapiro, J. Walter, D. Kozono, J. Ablain, M. Delord, J.-B. Lazaro and members of the D'Andrea laboratory for discussions. We also thank H. Kim for providing DNA constructs, L. Moldovan for providing GST tagged RAD51, S. Johnson for providing MDA-MB 436 cells, A. Y. Li for technical assistance, L. Moreau for chromosomal breakage analysis, and K. Mouw for critical reading of the manuscript. We thank H. Zhang for providing access to shRNA screening data shown in Extended Data Fig. 6i. We thank C. Cazaux and J.-S. Hoffmann for providing pcDNA Flag-tagged Pol θ . R.C. received support from the Philippe Foundation and is a recipient of the Ovarian Cancer Research Fellowship (OCRF). This work was supported by NIH grants P50CA168504 and R01HL52725 and by grants from OCRF and BCRF.

Author Contributions R.C. conceived the study, performed experiments, and wrote the manuscript. J.C.L. and T.Y. purified Pol θ fragments from insect cells and performed ATPase and gel shift assays. R.A. performed D-loop formation assays. I.H. and S.J.E. performed the DNA fibres assay. B.P. performed mice work and analysed *in vivo* data. M.I.R.P. and S.J.B. performed the Pol θ peptide array and the RAD51–ssDNA filament assembly and release assays. K.W.O. scored RAD51 foci. P.A.K. curated TCGA datasets for Figure 3a and Extended Data Figures 5h and 9e and provided clinical perspectives. A.D.D. conceived the study and wrote the manuscript. All authors approved the final version of the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.D.D. (alan_dandrea@dfci.harvard.edu).

METHODS

Bioinformatic analysis. Gene Set Enrichment Analysis algorithm (GSEA, <http://www.broadinstitute.org>) was performed for the data sets summarized in Supplementary Table 1. TransLesion Synthesis (TLS) and polymerase gene sets are described in Supplementary Table 3. Row expression data were downloaded from Gene Expression Omnibus (GEO). Quantile normalizations were performed using the RMA routine through GenePattern. GSEA was run using GenePattern (<http://www.broadinstitute.org>) and corresponding *P* values were computed using 2,000 permutations. The DNA repair gene set used in Extended Data Fig. 1g has been determined according to a list of 151 DNA genes previously used³¹. GSEA analysis for 151 repair genes has been performed on the ovarian serous data sets (GSE14001, GSE14007, GSE18520, GSE16708, GSE10971). The list of 20 genes shown in Extended Data Fig. 1g represents the top 20 expressed gene in cancer samples (median of the 5 data sets). The waterfall plot in Extended Data Fig. 1h was generated as follows: the 20 genes defined in Extended Data Fig. 1g were used as a gene set; GSEA for indicated data sets was performed and the nominal *P* values were plotted. Supervised analysis of gene expression for GSE9891 was performed with respect to differential expression that differentiated the third of tumours with highest *POLQ* expression from the two-thirds with lowest *POLQ* levels. A list of the 200 most differentially expressed probe sets between the 2 groups (Supplementary Table 2) with false discovery rate <0.05 was analysed for biological pathways (hypergeometrical test; <http://www.broadinstitute.org>). TCGA data sets were accessed through the public TCGA data portal (<https://tcga-data.nci.nih.gov/tcga/>). Fig. 3a reflects *POLQ* gene expression in the ovarian carcinoma data set GSE9891, uterine carcinoma TCGA and breast carcinoma TCGA. Normalization of *POLQ* expression values across data sets was performed using *z*-score transformation. *POLQ* expression values were subdivided in subgroups reflecting the stage of the disease (for GSE9891: grade 3 ovarian serous carcinoma, *n* = 143 compared to type 1 (grade 1) ovarian cancers, *n* = 20; for uterine: serous-like tumours, *n* = 60 compared to the rest of the tumours, *n* = 172; for breast: basal-like breast carcinoma, *n* = 80 compared to the rest of the tumours, *n* = 421). Progression-free survival curves were generated by the Kaplan–Meier method and differences between survival curves were assessed for statistical significance with the log-rank test. In the absence of a clinically defined cut-off point for *POLQ* expression levels, we divided patients into 2 groups: those with *POLQ* mRNA levels equal to or above the median (*POLQ* high group) and those with values below the median (*POLQ* low group). We then analysed the correlation of *POLQ* with outcome in each group. Patients with cyclin E1 (CCNE1) amplification (resistant to CDDP) were excluded from the analysis. For mutation count, we accessed data from tumours included in the TCGA data sets for which gene expression and whole-exome DNA sequencing was available. Data were accessed through the public TCGA data portal and the cBioPortal for Cancer Genomics (<http://www.cbioportal.org>). For each TCGA data set, non-synonymous mutation count was assessed in tumours with the highest *POLQ* expression (top 33%) and compared to tumours with low *POLQ* expression (the remaining 67%). In the uterine TCGA³⁰, we curated all tumours except the ultra and hyper-mutated group (that is, POLE and MSI tumours). In the breast TCGA³², all tumours were analysed. In the ovarian TCGA¹, we curated tumours harbouring molecular alterations (via mutation and epigenetic silencing) of the HR pathway.

Plasmid construction. To facilitate subcloning, a silent mutation (A390A) was introduced into the *POLQ* gene sequence to remove the unique XhoI cutting site. Full-length or truncated *POLQ* cDNA were PCR-amplified and subcloned into pcDNA3-N-Flag, pFastBac-C-Flag, pOZ-C-Flag-HA, or GFP-C1 vectors to generate the various constructs. Point mutations and loop deletions were introduced by QuikChange II XL Site-Directed Mutagenesis Kit (Agilent Technologies) and confirmed by DNA sequencing. For Polθ rescue experiments (Fig. 4g, h and Extended Data Fig. 3d, e), *POLQ* cDNA constructs resistant to siPolθ1 were generated into the pOZ-C-Flag-HA vector and the constructs were stably expressed in indicated cell line by retroviral transduction. The Polθ ATPase catalytically-dead mutant (A-dead) was generated by mutating the Walker A and B motifs (Walker A: K121A and Walker B: D216A, E217A). pOZ-C-Flag-HA Polθ constructs were generated for retroviral transduction, and stable cells were selected using magnetic Dynabeads (Life Technologies) conjugated to the IL2R antibody (Millipore).

siRNA and shRNA sequence information. For siRNA-mediated knockdown, the following target sequences were used: *POLQ* (Qiagen POLQ_1 used as siPolθ1 and Qiagen POLQ_6 used as siPolθ2); *BRCA1* (Qiagen BRCA1_13); *PARP1* (Qiagen PARP1_6); *REV1* (5'-CAGCGCAUCUGGCCAAAGAA-TT-3'); *BRCA2* (5'-GAAGAAUGCAGGUUAAUATT-3'); *BLM* (5'-AUCAGCUAGAGGCGAUAATT-3'); *FANCD2* (5'-GGAGAUUGAUGUCUACUATT-3') and *PARI* (5'-AGGACACAUUGAAAGGAUUGUCUATT-3'). AllStars negative control siRNA (Qiagen) served as the negative control. ShRNAs targeting human *FANCD2* was previously generated in the pTRIP/DU3-MND-GFP vector³³. ShRNAs targeting human *POLQ* (CGGGCCTCTTTAGATATAAAT), human *BRCA2* (AAGAAGAATGCAGGTTTAATA) or control (Scr, scramble) were generated in the pLKO-1

vector. *POLQ* (V2THS_198349) and non-silencing TRIPZ-RFP doxycycline-inducible shRNA were purchased from Open Biosystems. All shRNAs were transduced using lentivirus.

Immunoblot analysis, fractionation and pull-down assays. Cells were lysed with 1% NP-40 lysis buffer (1% NP-40, 300 mM NaCl, 0.1 mM EDTA, 50 mM Tris (pH 7.5)) supplemented with protease inhibitor cocktail (Roche), resolved by NuPAGE (Invitrogen) gels, and transferred onto nitrocellulose membrane, followed by detection using the LAS-4000 Imaging system (GE Healthcare Life Sciences). For immunoprecipitation, cells were lysed with 300 mM NaCl lysis buffer, and the lysates were diluted to 150 mM NaCl before immunoprecipitation. Lysates were incubated with anti-Flag agarose resin (Sigma) followed by washes with 150 mM NaCl buffer. *In vitro* transcription and translation reactions were carried out using the TNT T7 Quick Coupled Transcription-Translation System (Promega). For cellular fractionation, cells were incubated with low-salt permeabilization buffer (10 mM Tris (pH 7.3), 10 mM KCl, 1.5 mM MgCl₂) with protease inhibitor on ice for 20 min. Following centrifugation, nuclei were resuspended in 0.2 M HCl and the soluble fraction was neutralized with 1 M Tris-HCl (pH 8.0). Nuclei were lysed in 150 mM NaCl and following centrifugation, the chromatin pellet was digested by micrococcal nuclease (Roche) for 5 min at room temperature. Recombinant GST–RAD51 and GST–PCNA fusion protein were expressed in BL21 strain and purified using glutathione-Sepharose beads (GE Healthcare) as previously described¹⁵. Beads with equal amount of GST or GST–RAD51 were incubated with *in vitro* translated Flag-tagged Polθ variants in 150 mM NaCl lysis buffer.

Antibodies and chemicals. Antibodies used in this study included: anti-PCNA (PC-10), anti-FANCD2 (FI-17), anti-RAD51 (H-92), anti-GST (B14), and histone H3 (FL-136) and anti-vinculin (H-10) (Santa Cruz); anti-Flag (M2) (Sigma); anti-pS317CHK1 (2344), anti-pT68CHK2 (2661) (Cell signalling); anti-pS824KAP-1 (A300-767A) (Bethyl); anti-pS317γH2AX (05636) (Millipore); anti-pS15p53 (ab1431) and anti-Polθ (ab80906) (abcam); anti-BrdU (555627) (BD Pharmingen). Mitomycin C (MMC), cis-diamminedichloroplatinum(II) (cisplatin, CDDP), and hydroxyurea (HU) were purchased from Sigma. The PARPi rucaparib (AG-014699) was purchased from Selleckchem and ABT-888 from AbbVie. Rucaparib was used for all *in vitro* assays and ABT-888 was used for all *in vivo* experiments.

Chromosomal breakage analysis. 293T and VU 423 cells were twice-transfected with siRNAs for 48 h and incubated for 48 h with or without the indicated concentrations of MMC. For complementation studies on 293T shFANCD2, *POLQ* cDNA constructs were transfected 24 h after the first siRNA transfection. Cells were exposed for 2 h to 100 ng ml⁻¹ of colcemid and treated with a hypotonic solution (0.075 M KCl) for 20 min and fixed with 3:1 methanol/acetic acid. Slides were stained with Wright's stain and 50 metaphase spreads were scored for aberrations. The relative number of chromosomal breaks was calculated relative to control cells (si Scr). Radial figures were excluded from the analysis for clarity in Fig. 4b.

Reporter assays and immunofluorescence. HR and alt-EJ efficiency was measured using the DR-GFP (HR efficiency) and the alt-EJ reporter assay, performed as previously described^{14,27,34}. Briefly, 48 h before transfection of Scl cDNA, U2OS-DR-GFP cells were transfected with the indicated siRNA or PARPi (1 μM). The HR activity was determined by FACS quantification of viable GFP-positive cells 96 h after Scl was transfected. For RAD51 immunofluorescence experiments, cells were transfected with indicated siRNA 48 h before treatment with HU (2 mM) or IR (10 Gy). For complementation studies, Polθ cDNA constructs were either transfected 24 h after siRNA transfection (Fig. 2b, c and Extended Data Fig. 9b) or stably expressed in the indicated cell line (Extended Data Fig. 3d, e). 6 h after HU or IR treatment, cells were fixed with 4% paraformaldehyde for 10 min at room temperature, followed by extraction with 0.3% Triton X-100 for 10 min on ice. Antibody staining was performed at room temperature for 1 h. For quantification of RAD51 foci in BrdU positive cells, cells were transfected with indicated siRNA 48 h before treatment with IR (10 Gy). Then 2 h after IR treatment, cells were treated with BrdU pulse (10 μM) for 2 h and subsequently fixed with 4% paraformaldehyde and stained for RAD51 as described above. Cells were then fixed in ethanol (4 °C, overnight), treated with 1.5 M HCl for 30 min and stained for BrdU antibody. The relative number of cells with more than 10 RAD51 foci was calculated relative to control cells (si Scr). Statistical differences between cells transfected with siRNAs (si Polθ1, si Polθ2, si BRCA2, si PARI or si BLM relative to control (si Scr) were assessed. For GFP fluorescence, cells were grown on coverslips, treated with UV (24 h after GFP–Polθ transfection; 20 J m⁻²), fixed with 4% paraformaldehyde for 10 min at 25 °C 4 h after the UV treatment, washed three times with PBS and mounted with DAPI-containing mounting medium (Vector Laboratories). When indicated, cells were treated with PARPi (1 μM) 24 h before GFP–Polθ transfection. Images were captured using a Zeiss AX10 fluorescence microscope and AxioVision software. Cells with GFP foci were quantified by counting number of cells with more than five foci. At least 150 cells were counted for each sample.

Cell survival assays. For assessing cellular cytotoxicity, cells were seeded into 96-well plates at a density of 1,000 cells per well. Cytotoxic drugs were serially diluted

in media and added to the wells. At 72 h, CellTiter-Glo reagent (Promega) was added to the wells and the plates were scanned using a luminescence microplate reader. Survival at each drug concentration was plotted as a percentage of the survival in drug-free media. Each data point on the graph represents the average of three measurements, and the error bars represent the standard deviation. For clonogenic survival, 1,000 cells per well were seeded into 6-well plates and treated with cytotoxic drugs the next day. For MMC and PARPi, cells were treated continuously with indicated drug concentrations. For CDDP, cells were treated for 24 h and cultured for 14 days in drug-free media. Colony formation was scored 14 days after treatment using 0.5% (w/v) crystal violet in methanol. Survival curves were expressed as a percentage \pm s.e.m. over three independent experiments of colonies formed relative to the DMSO-treated control.

Cell cycle analysis. A2780 cells expressing Scr or Polθ shRNA were synchronized by a double thymidine block (Sigma) and subsequently exposed to MMC (1 $\mu\text{g ml}^{-1}$ for 2 h), IR (10 Gy) or HU (2 mM, overnight). At the indicated time points following drug release, cells were fixed in chilled 70% ethanol, stored overnight at -20°C , washed with PBS, and resuspended in propidium iodide. A fraction of those cells was analysed by immunoblotting for DNA damage response proteins. The immunoblot analysis of γH2AX shows staining after 0, 24, 48 and 72 h of HU treatment. For proliferation experiments, cells were incubated with 5-ethynyl-2'-deoxyuridine (EdU) (10 μM) for 1 h at each time point after MMC exposure (1 $\mu\text{g ml}^{-1}$ for 2 h). Cells were washed and resuspended in culture medium for 2 h before being analysed by flow cytometry. Edu staining was performed using the Click-iT Edu kit (Life Technologies).

DNA fibre analysis. A2780 cells expressing Scr or Polθ shRNA were incubated with 25 μM chlorodeoxyuridine (CldU) (Sigma, C6891) for 20 min. Cells were then treated with 2 mM hydroxyurea (HU) for 2 h and incubated in 250 μM iododeoxyuridine (IdU) (Sigma, I7125) for 25 min after washout of the drug. Spreading of DNA fibres on glass slides was done as previously reported¹⁹. Glass slides were then washed in distilled water and in 2.5 M HCl for 80 min followed by three washes in PBS. The slides were incubated for 1 h in blocking buffer (PBS with 1% BSA and 0.1% NP-40) and then for 2 h in rat anti-BrdU antibody (1:250, Abcam, ab6326). After washing with blocking buffer, the slides were incubated for 2 h in goat anti-rat Alexa 488 antibody (1:1,000, Life Technologies, A-11006). The slides were then washed with PBS and 0.1% NP-40 and then incubated for 2 h with mouse anti-BrdU antibody diluted in blocking buffer (1:100, BD Biosciences, 347580). Following an additional wash with PBS and 0.1% NP-40, the fibres were stained for 2 h with chicken anti-mouse Alexa 594 (1:1,000, Life Technologies, A-21201). At least 150 fibres were counted per condition. Pictures were taken with an Olympus confocal microscope and the fibres were analysed by ImageJ software. The number of stalled or collapsed forks were measured by DNA fibres that had incorporated only CldU. Stalled or collapsed forks counted in Polθ-depleted cells is expressed as fold-change after HU treatment relative to the fold-change observed in control cells, which was arbitrarily set to 1.

SupF mutagenesis assay. 293T cells twice-transfected with siRNAs for 48 h were then transfected with undamaged or damaged (UVC, 1,000 J m^{-2}) pSP189 plasmids using GeneJuice (Novagen). After 48 h, plasmid DNA was isolated with a miniprep kit (Promega) and digested with DpnI. After ethanol precipitation, extracted plasmids were transformed into the β -galactosidase—MBM7070 indicator strain through electroporation (GenePulser X Cell; Bio-Rad) and plated onto LB plates containing 1 mM IPTG, 100 $\mu\text{g ml}^{-1}$ 5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside and 100 $\mu\text{g ml}^{-1}$ ampicillin. White and blue colonies were scored using ImageJ software, and the mutation frequency was calculated as the ratio of white (mutant) to total (white plus blue) colonies.

POLQ gene expression. RNA samples extracted using the TRIzol reagent (Invitrogen) were reverse transcribed using the Transcriptor Reverse Transcriptase kit (Roche) and oligo dT primers. The resulting cDNA was used to analyse POLQ expression by RT-qPCR using with QuantiTect SYBRGreen (Qiagen), in an iCycler machine (Bio-Rad). POLQ gene expression values were normalized to expression of the housekeeping gene *GAPDH*, using the ΔCt method and are shown on a log₂ scale. The primers used for POLQ are as follows: POLQ primer 1 (forward: 5'-TATCTG CTGGAACCTTTTGCTGA-3'; reverse: 5'-CTCACACCATTTCTTTGATGGA-3'); POLQ primer 2 (forward: 5'-CTACAAGTGAAGGGAGATGAGG-3'; reverse: 5'-TCAGAGGGTTTACCAATCC-3').

Polθ purification from insect SF9 cells. A Polθ fragment (ΔPol2) containing the ATPase domain with a RAD51 binding site (amino acids 1 to 1,000) was cloned into pFastBac-C-Flag and purified from baculovirus-infected SF9 insect cells as previously described³⁵. Briefly, SF9 cells were seeded in 15-cm dishes at 80–90% confluency and infected with baculovirus. Three days post-infection, cells were collected and lysed in 500 mM NaCl lysis buffer (500 mM NaCl, 0.01% NP-40, 0.2 mM EDTA, 20% glycerol, 1 mM DTT, 0.2 mM PMSF, 20 mM Tris (pH 7.6)) supplemented with Halt protease inhibitor cocktail (Thermo Scientific) and calpain I inhibitor (Roche) and the protein was eluted in lysis buffer supplemented with

0.2 mg ml⁻¹ of Flag peptide (Sigma). The protein was concentrated in lysis buffer using 10 kDa centrifugal filters (Amicon). The protein was quantified by comparing its staining intensity (Coomassie-R250) with that of BSA standards in a 7% Tris-glycine SDS-PAGE gel. Purified protein was flash-frozen in small aliquots in liquid nitrogen and stored at -80°C .

Radiometric ATPase assay. Each 10 μl reaction consisted of 200 nM ATP, reaction buffer (20 mM Tris-HCl (pH 7.6), 5 mM MgCl₂, 0.05 mg ml⁻¹ BSA, 1 mM DTT), and 5 μCi of [γ -³²P]ATP. For corresponding reactions, ssDNA, dsDNA, and forked DNA were added to the reaction in excess at a final concentration of 600 nM. Once all of the non-enzymatic reagents were combined, recombinant Polθ was added to start the ATPase reaction. After incubation for 90 min at room temperature, stop buffer (125 mM EDTA (pH 8.0)) was added and approximately ~ 0.05 μCi was spotted onto PEI-coated thin-layer chromatography (TLC) plates (Sigma). Unhydrolyzed [γ -³²P]ATP was separated from the released inorganic phosphate [³²P]_i with 1 M acetic acid, 0.25 M lithium chloride as the mobile phase. TLC plates were exposed to a phosphor screen and imaged with the BioRad Imager PMC. ssDNA, dsDNA, and forked DNA were generated as previously described³⁵. To remove any contaminating ssDNA, dsDNA and forked DNA were gel purified after annealing. Spots corresponding to [γ -³²P]ATP and the released inorganic phosphate [³²P]_i were quantified (in units of pixel intensity) and the fraction of ATP hydrolysed calculated for each Polθ concentration.

Electrophoretic mobility gel shift assay (EMSA). Binding of Polθ to ssDNA was assessed using EMSA. 60-mer single-stranded DNA (ssDNA) or double-stranded DNA (dsDNA) oligonucleotides (5 nM) were incubated with increasing amount of Polθ (0, 5, 10, 50, or 100 nM) in 10 μl of binding buffer (20 mM HEPES-K⁺, (pH 7.6), 5 mM magnesium acetate, 0.1 $\mu\text{g ml}^{-1}$ BSA, 5% glycerol, 1 mM DTT, 0.2 mM EDTA, and 0.01% NP-40) for 1 h on ice. Polθ protein was added at a tenfold dilution so that the final salt concentration was approximately 50 mM NaCl. The ssDNA probes were 5' fluorescently labelled with IRDye-700 (IDT). After incubation, the samples were analysed on a 5% native polyacrylamide/0.5X TBE gel at 4°C . A fluorescent imager (Li-Cor) was used to visualize the samples in the gel.

RAD51 purification. Human GST-RAD51 was purified from bacteria as described³⁶. *Xenopus* Rad51 (xRad51) was purified as follows. N-terminally His-tagged SUMO-Rad51 was expressed in BL21 pLysS cells. Three hours after induction with 1 mM IPTG, cells were collected and resuspended in buffer A (50 mM Tris-Cl (pH 7.5), 350 mM NaCl, 25% sucrose, 5 mM β -mercaptoethanol, 1 mM PMSF and 10 mM imidazole). Cells were lysed by supplementation with Triton X-100 (0.2% final concentration), three freeze-thaw cycles and sonication (20 pulses at 40% efficiency). The soluble fraction was separated by centrifugation and incubated with 2 ml of Ni-NTA resin (Qiagen) for 1 h at 4°C . After washing the resin with 100 ml of wash buffer (buffer A supplemented with 1 M NaCl, final concentration), the salt concentration was brought down to 350 mM. His-SUMO-Rad51 was eluted with a linear gradient of imidazole from 10–300 mM in buffer A. Eluted fractions were analysed by SDS-PAGE. His-SUMO-Rad51 containing fractions were pooled and supplemented with Ulp1 protease to cleave the His-SUMO tag and dialysed overnight into buffer B (50 mM Tris-Cl (pH 7.5), 350 mM NaCl, 25% sucrose, 10% glycerol, 5 mM β -mercaptoethanol, 10 mM imidazole and 0.05% Triton X-100). The dialysed fraction was incubated with Ni-NTA resin for 1 h at 4°C and the Rad51 containing flow-through fraction was collected and dialysed overnight into buffer C (100 mM potassium phosphate (pH 6.8), 150 mM NaCl, 10% glycerol, 0.5 mM DTT and 0.01% Triton-X). Rad51 was further purified by hydroxyapatite (Bio-Rad) chromatography. After washing with ten column volumes of buffer C, Rad51 was eluted with a linear gradient of potassium phosphate (pH 6.8) from 100–800 mM. Rad51 containing fractions were analysed by SDS-PAGE and dialysed into storage buffer (20 mM HEPES-KOH (pH 7.4), 150 mM NaCl, 10% glycerol, 0.5 mM DTT). Purified protein was flash-frozen in small aliquots in liquid nitrogen and stored at -80°C .

D-loop assay. D-loop formation assays were performed using xRad51 and conducted as previously described³⁷. Briefly, nucleofilaments were first formed by incubating RAD51 (1 μM) with end-labelled 90-mer ssDNA (3 $\mu\text{M nt}$) at 37°C for 10 min in reaction buffer containing 20 mM HEPES-KOH (pH 7.4), 1 mM ATP, 1 mM MgCl₂, 1 mM DTT, BSA (100 $\mu\text{g ml}^{-1}$), 20 mM phosphocreatine and creatine phosphokinase (20 $\mu\text{g ml}^{-1}$). After the 10 min incubation, increasing amounts of Polθ (0, 0.1, 0.5, or 1.0 μM) and RPA (200 nM) were added and incubated for an additional 15 min at 37°C . The reaction was then supplemented with 1 mM CaCl₂ followed by further incubation at 37°C for 15 min. D-loop formation was initiated by the addition of supercoiled dsDNA (pBS-KS (–), 79 $\mu\text{M bp}$) and incubation at 37°C for 15 min. D-loops were analysed by electrophoresis on a 0.9% agarose gel after deproteinization. Gel was dried and exposed to a PhosphorImager (GE Healthcare) screen for quantification.

Substitution peptide arrays and RAD51-ssDNA filament experiments. Substitution peptide arrays were performed as previously described¹⁷. RAD51 displacement assays were performed as follows. Binding reactions (10 μl) contained 5'-³²P-end-labelled

DNA substrates (0.5 ng of 60 mer ssDNA) and various amounts of human RAD51 and/or Polθ in binding buffer (40 mM Tris-HCl (pH 7.5), 50 mM NaCl, 10 mM KCl, 2 mM DTT, 5 mM ATP, 5 mM MgCl₂, 1 mM DTT, 100 mg ml⁻¹ BSA) were conducted at room temperature. After 5 min incubation with Polθ and a further 5 min incubation with RAD51 or vice versa, an equimolar amount of cold DNA substrate was added to the reaction. Products were then analysed by electrophoresis through 10% PAGE (200 V for 40 min in 0.5× Tris-borate-EDTA buffer) and visualized by autoradiography.

Interbreeding of the *Fancd2* and *Polq* mice. For the characterization of *Fancd2/Polq* conditional knockouts, we crossed C57BL/6J mice (Jackson Laboratory). *Fancd2*^{+/-}*Polq*^{+/-} mice, previously generated in our laboratory²², were crossed with *Fancd2*^{+/-}*Polq*^{+/-} mice² to generate *Fancd2*^{+/-}*Polq*^{+/-} mice. These double heterozygous mice were then interbred, and the offspring from these mating pairs were genotyped using PCR primers for *Fancd2* and *Polq*. A statistical comparison of the observed with the predicted genotypes was performed using a two-sided Fisher's exact test. Primary MEFs were generated from E13.5 to E15 embryos and cultured in RPMI supplemented with 15% fetal bovine serum and 1% penicillin-streptomycin. All data generated in the study were extracted from experiments performed on primary MEFs from passage 1 to passage 4. The primers used for mice genotyping are as follows: *Fancd2* PCR primers OST2cF (5'-CATGCATATAGGAACCCGAAGG-3'), OST2aR (5'-CAGGACCTTTGGAGAAGCAG-3') and LTR2bF (5'-GCGTTACTTAAGCTAGCTTG-3'); *Polq* PCR primers IMR5973 (5'-TGCAGTG TACAGATGTTACTTTT-3'), IMR 5974 (5'-TGGAGGTAGCATTTCTTCTC-3'), IMR 5975 (5'-TCACTAGGTTGGGGTTCTC-3') and IMR 5976 (5'-CATCAGAAGCTGACTCTAGAG-3'). Specific PCR conditions are available upon request.

Studies of xenograft-bearing CrTac:NCr-Foxn1nu mice. The Animal Resource Facility at The Dana-Farber Cancer Institute approved all housing situations, treatments and experiments using mice. No more than five mice were housed per air-filtered cage with *ad libitum* access to standard diet and water, and were maintained in a temperature- and light-controlled animal facility under pathogen-free conditions. All mice described in this text were drug and procedure naive before the start of the experiments. For every xenograft study, we subcutaneously implanted approximately 1.0×10^6 A2780 cells (1:1 in Matrigel Matrix, BD Biosciences) into both flanks of 6–8-week-old female CrTac:NCr-Foxn1nu mice (Taconic). Doxycycline (Sigma) was added to the food (625 p.p.m.) and bi-weekly (Tuesday and Friday) to the water (200 µg ml⁻¹) for mice bearing tumours that reached 100–200 mm³. Roughly one week (5–6 days) after the addition of doxycycline to the diet, mice were randomized to twice daily treatment schedules with vehicle (0.9% NaCl) or PARPi (ABT-888; 50 mg per kg body weight) by oral gavage administration for the indicated number of weeks. Overall survival was determined using Kaplan-Meier analyses performed with log-rank tests to assess differences in median survival for each shRNA condition (shScr or shPolθ) and each treatment condition (vehicle or PARPi) (GraphPad Prism 6 Software). For competition assays, A2780 cells expressing FANCD2-GFP shRNA (GFP positive cells) or a combination of FANCD2-GFP shRNA with (doxycycline inducible) Scr-RFP or Polθ-RFP shRNA (GFP-RFP positive cells) were mixed at an equal ratio of GFP to GFP-RFP positive cells, and thereafter injected into nude mice given doxycycline-containing diets and treated with either vehicle or PARPi or CDDP. For competition assays, mice received identical doxycycline and PARPi drug treatment. For the cisplatin

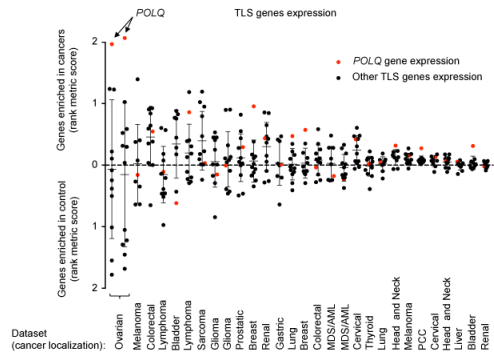
competition assay, mice were randomized into semi-weekly treatment regimens with vehicle (0.9% NaCl) or CDDP (5 mg per kg body weight) by intraperitoneal injection. After three to four weeks of treatment, mice were euthanized and tumours were grown *in vitro*, in the presence of doxycycline (2 µg ml⁻¹ for 4 days). The relative ratio of GFP to GFP-RFP positive cells was determined by FACS analysis. Tumour volumes were calculated bi-weekly using caliper measurements (length × width²)/2. Growth curves were plotted as the mean tumour volume (mm³) for each treatment group; relative tumour volume (RTV) indicates the change in tumour volume at a given time point relative to the tumour volume at the day of initial measurement (volume of approximately 0.15 cm³) which was arbitrarily set to 1. Mice were unbiasedly assigned into different treatment groups. Drug treatment and outcome assessment was performed in a blinded manner. Mice were monitored every day and euthanized by CO₂ inhalation when tumour size (≥ 2 cm), tumour status (necrosis/ulceration) or body weight loss (≥ 20%) reached ethical endpoint, according to the rules of the Animal Resource Facility at The Dana-Farber Cancer Institute.

Immunohistochemical staining. We stained formalin-fixed paraffin-embedded sections of harvested xenografts with antibodies specific for γH2AX (pSer139) (Upstate Biotechnology) and Ki67 (Dako). At least two xenografts were scored for each treatment. Tumours were collected 3 weeks after treatment. At least five 40 × fields were scored. The mean ± s.e.m. percentage of positive cells from five images in each treatment group was calculated.

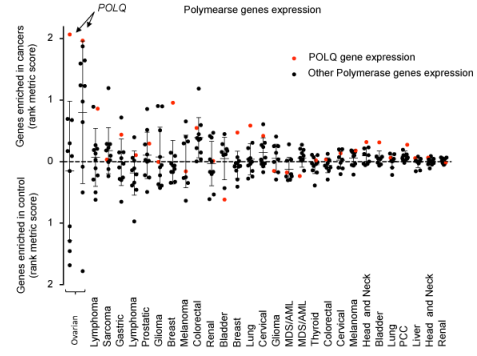
Statistical analysis. Unless stated otherwise, all data are represented as mean ± s.e.m. over at least three independent experiments, and significance was calculated using the Student's *t*-test. Asterisks indicate statistically significant (**P* < 0.05; ***P* < 10⁻²; ****P* < 10⁻³) values. All the *in vivo* experiments were run with at least 6 tumours from 6 mice for each condition. No statistical methods were used to predetermine sample size.

31. Kang, J., D'Andrea, A. D. & Kozono, D. A DNA repair pathway-focused score for prediction of outcomes in ovarian cancer treated with platinum-based chemotherapy. *J. Natl. Cancer Inst.* **104**, 670–681 (2012).
32. The Cancer Genome Atlas Network Comprehensive molecular portraits of human breast tumours. *Nature* **490**, 61–70 (2012).
33. Ceccaldi, R. *et al.* Bone marrow failure in Fanconi anemia is triggered by an exacerbated p53/p21 DNA damage response that impairs hematopoietic stem and progenitor cells. *Cell Stem Cell* **11**, 36–49 (2012).
34. Bennardo, N., Cheng, A., Huang, N. & Stark, J. M. Alternative-NHEJ is a mechanistically distinct pathway of mammalian chromosome break repair. *PLoS Genet.* **4**, e1000110 (2008).
35. Yusufzai, T. & Kadonaga, J. T. HARP is an ATP-driven annealing helicase. *Science* **322**, 748–750 (2008).
36. Barber, L. J. *et al.* RTEL1 maintains genomic stability by suppressing homologous recombination. *Cell* **135**, 261–271 (2008).
37. Amunugama, R., Groden, J. & Fishel, R. The HsRAD51B-HsRAD51C stabilizes the HsRAD51 nucleoprotein filament. *DNA Repair* **12**, 723–732 (2013).
38. Pitroda, S. P. *et al.* DNA repair pathway gene expression score correlates with repair proficiency and tumor sensitivity to chemotherapy. *Sci. Transl. Med.* **6**, 229ra242 (2014).
39. Aggarwal, M., Sommers, J. A., Shoemaker, R. H. & Brosh, R. M., Jr. Inhibition of helicase activity by a small molecule impairs Werner syndrome helicase (WRN) function in the cellular response to DNA damage or replication stress. *Proc. Natl. Acad. Sci. USA* **108**, 1525–1530 (2011).

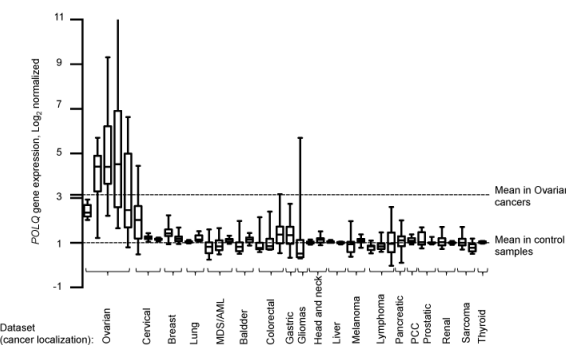
a



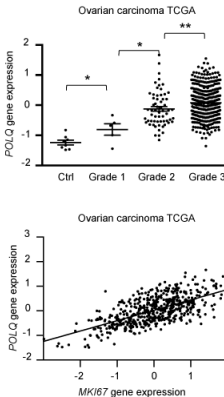
b



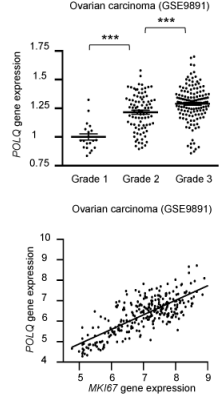
c



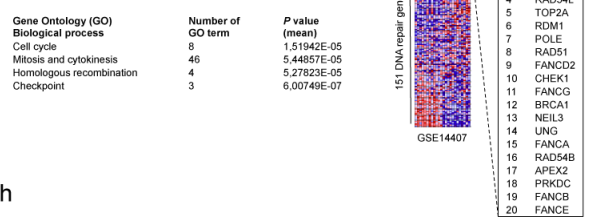
d



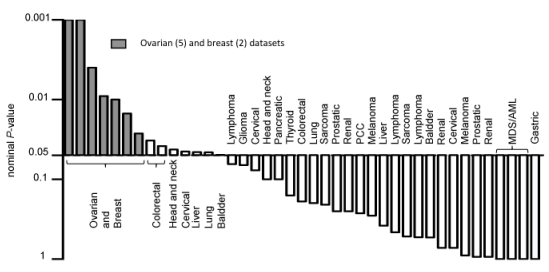
e



f



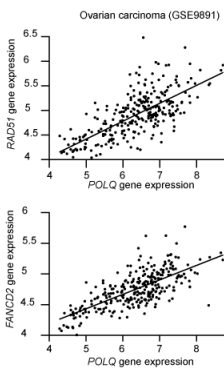
h



k

Gene Ontology (GO) molecular function	GO ID	P value
DNA helicase activity	GO:0003678	0.00002
helicase activity	GO:0004386	0.0001
nucleoside-triphosphatase activity	GO:0017111	0.0002
pyrophosphatase activity	GO:0016462	0.0002
hydrolase activity, acting on acid anhydrides	GO:0016817	0.0002
hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides	GO:0016818	0.0002
catalytic activity	GO:0003624	0.0003
DNA-directed DNA polymerase activity	GO:0003887	0.0003
DNA polymerase activity	GO:0004061	0.0005
damaged DNA binding	GO:0003694	0.0009

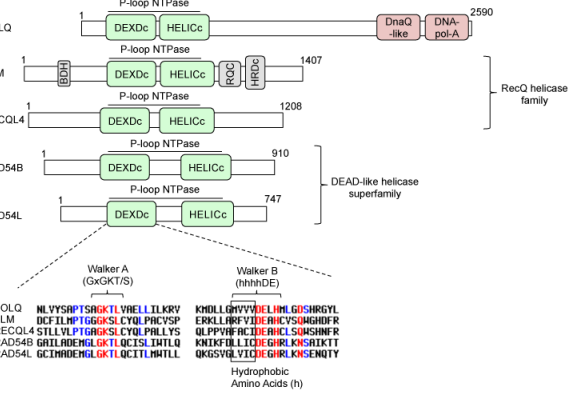
i



j

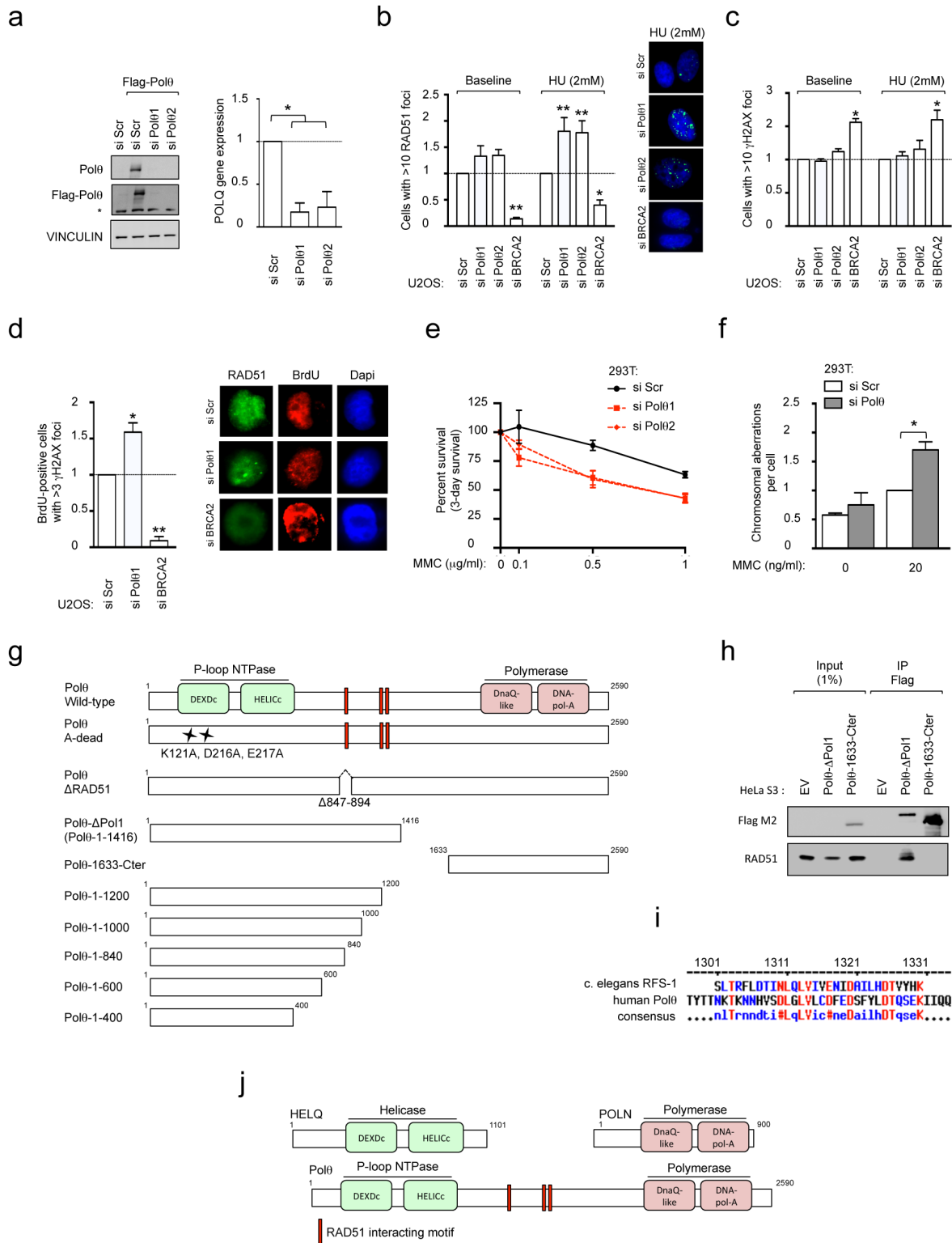
POLQ	DNA repair activity
TOPBP1	yes
BLM	yes
RAD54L	yes
PLK4	no
FANCI	yes
SMC4	no
BRCA1	yes
FANCD2	yes
ATAD5	yes
TRAP1	no

l



Extended Data Figure 1 | *POLQ* is highly expressed in epithelial ovarian cancers (EOCs) and *POLQ* expression correlates with expression of HR genes. **a, b,** Gene set enrichment analysis (GSEA) for expression of TransLesion Synthesis (TLS) (**a**) and polymerase (**b**) genes between primary cancers and control samples in 28 independent data sets from 19 different cancers types. Enrichment values (represented as a single dot for each gene in a defined data set) were determined using the rank metric score to compare expression values between cancers and control samples. Dots above the dashed line reflect enrichment in cancer samples, whereas dots below the dashed line show gene expression enriched in control samples. Data sets were ranked based on the amplitude of the rank metric score and plotted as shown. **c,** *POLQ* gene expression in 40 independent data sets from 19 different cancer types. For each data set, *POLQ* values were expressed as fold-change differences relative to the mean expression in control samples, which was arbitrarily set to 1. **d,** *POLQ* expression correlates with tumour grade and MKi67 gene expression in the ovarian TCGA ($n = 494$ patients with ovarian carcinoma (grade 1, $n = 5$; grade 2, $n = 61$; grade 3, $n = 428$) and control samples, $n = 8$). **e,** *POLQ* expression correlates with tumour grade MKi67 gene expression in the ovarian data set GSE9891 ($n = 251$ patients with ovarian serous and endometrial carcinoma for which grade status was available (grade 1, $n = 20$; grade 2, $n = 88$; grade 3, $n = 143$)). Statistical correlation was assessed using the Pearson test (for **d**: $r = 0.65$, $P < 10^{-3}$; for **e**: $r = 0.77$, $P < 10^{-3}$). **f,** Top-ranked biological pathways differentially expressed between samples expressing high levels of *POLQ* (high *POLQ*, first 33%, $n = 95$) relative to samples with low *POLQ* expression (low *POLQ*, 67%, $n = 190$) on the ovarian data set GSE9891 ($n = 285$ patients with ovarian carcinoma). Significance values were determined by the hypergeometrical test using the 200 most differentially expressed probe sets between the 2 groups (high *POLQ* and low *POLQ*). **g,** GSEA for expression of DNA repair genes between primary cancers

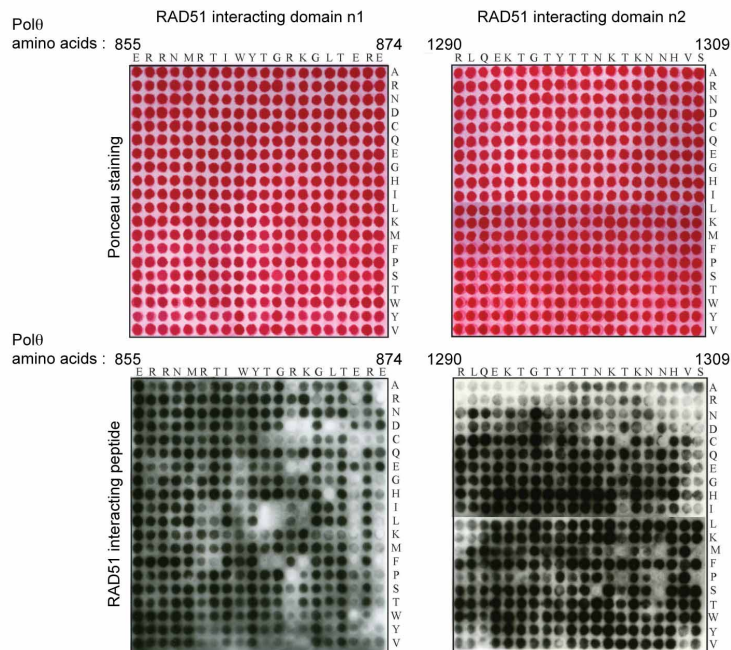
and control samples in 5 independent ovarian cancer data sets. A representative heat map showing differential gene expression between ovarian cancers and controls is shown from GSE14407. For each data set, DNA repair genes were ranked based on the metric score reflecting their enrichment in cancer samples. The top 20 DNA repair genes primarily expressed in cancer samples compared to control samples is shown on the right. **h,** GSEA for the top 20 DNA repair genes defined in **g** between primary cancers and control samples in 40 independent cancer data sets. The nominal P value was used as a measure of the expression enrichment in cancer samples and represented as a waterfall plot. When the gene set expression was enriched in control samples, the P value was arbitrarily set to 1. **i,** *POLQ* expression correlates with *RAD51* and *FANCD2* gene expression in 285 samples from the ovarian data set GSE9891. Statistical correlation was assessed using the Pearson test ($r = 0.71$, $P < 10^{-3}$). **j,** Top 10 genes that most closely correlated with *POLQ* expression (gene neighbours analysis) for 1,046 cell lines from the CCLE collection. DNA repair activity for these genes is indicated in the table. Increased HR gene expression is known to positively correlate with improved response to platinum based chemotherapy (a surrogate of HR deficiency) and thus can be predictive of decreased HR activity^{31,38}. Conceptually, a state of HR deficiency may lead to compensatory increased expression of other HR genes. **k,** Top-ranked Gene Ontology (GO) terms for the molecular functions encoded by the top 20 DNA repair genes defined in Extended Data Fig. 1g. **l,** Schematic representation of Polθ domain structure with the helicases (BLM, RECQL4, RAD54B and RAD54L) that co-expressed with Polθ (from Extended Data Fig. 1g). Conserved amino-acid sequences of ATP binding and hydrolysis motifs (namely Walker A and B) are indicated. Cox plots in **c** show twenty-fifth to seventy-fifth percentiles, with lines indicating the median, and whiskers indicating the smallest and largest values. For **d** and **e** (top panels), each dot represents the expression value from one patient, brackets show mean \pm s.e.m.



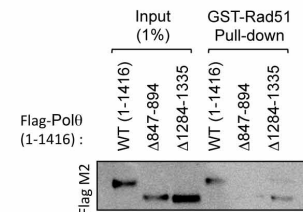
Extended Data Figure 2 | Polθ is a RAD51-interacting protein required for maintenance of genomic stability. **a**, siRNA sequences (siPolθ1 and siPolθ2) efficiently downregulate exogenously transfected Polθ protein. Polθ levels were detected by immunoblotting with Flag or Polθ antibody (left) and by RT-qPCR using 2 different sets of *POLQ* primers (right). The asterisk on the immunoblot indicates a non-specific band. Expression was normalized using *GAPDH* as a reference gene. *POLQ* gene expression values are displayed as fold-change differences relative to the mean expression in control cells, which was arbitrarily set to 1. **b**, Quantification of baseline and HU-induced RAD51 foci in U2OS cells transfected with the indicated siRNA. **c**, Quantification of baseline and HU-induced γH2AX foci in U2OS cells transfected with the

indicated siRNA. **d**, Quantification of IR-induced RAD51 foci in BrdU-positive U2OS cells transfected with the indicated siRNA. **e**, Polθ inhibition by siRNA induced a decrease in the cellular survival of 293T cells treated with MMC in a 3-day survival assay. **f**, Quantification of chromosomal aberrations in 293T cells transfected with the indicated siRNA. **g**, Schematic representation of Polθ truncation proteins used for RAD51 interaction studies. **h**, Endogenous RAD51 co-precipitates with Flag-tagged Polθ-ΔPol1 (Polθ-1-1,416) but not Polθ-1633-Cter, each stably expressed in HeLa cells. **i**, Sequence alignment between the RAD51-interacting motifs of *C. elegans* RFS-1 and human Polθ. **j**, Schematic of POLN domain structure with its homologues HELQ and POLN. All data show mean ± s.e.m.

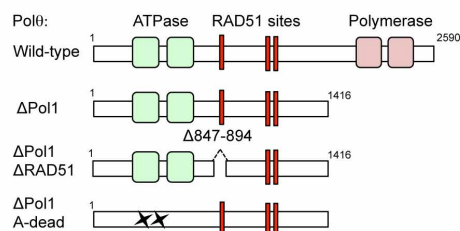
a



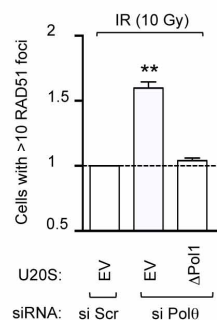
b



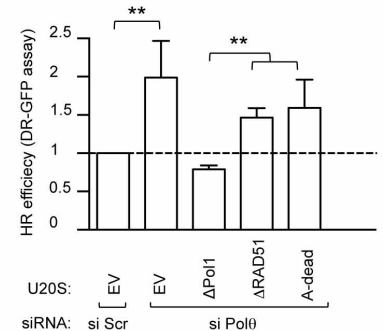
c



d

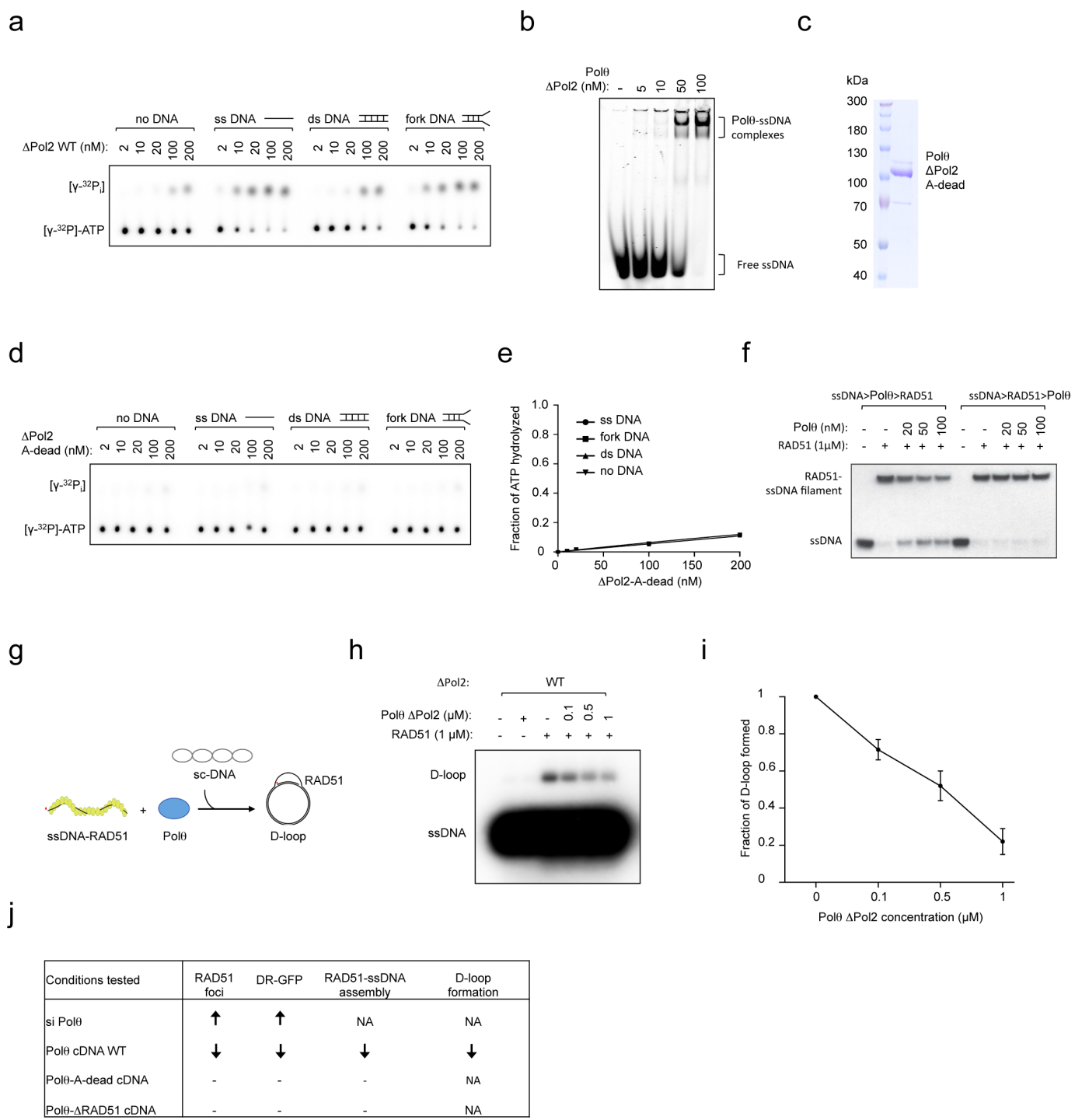


e

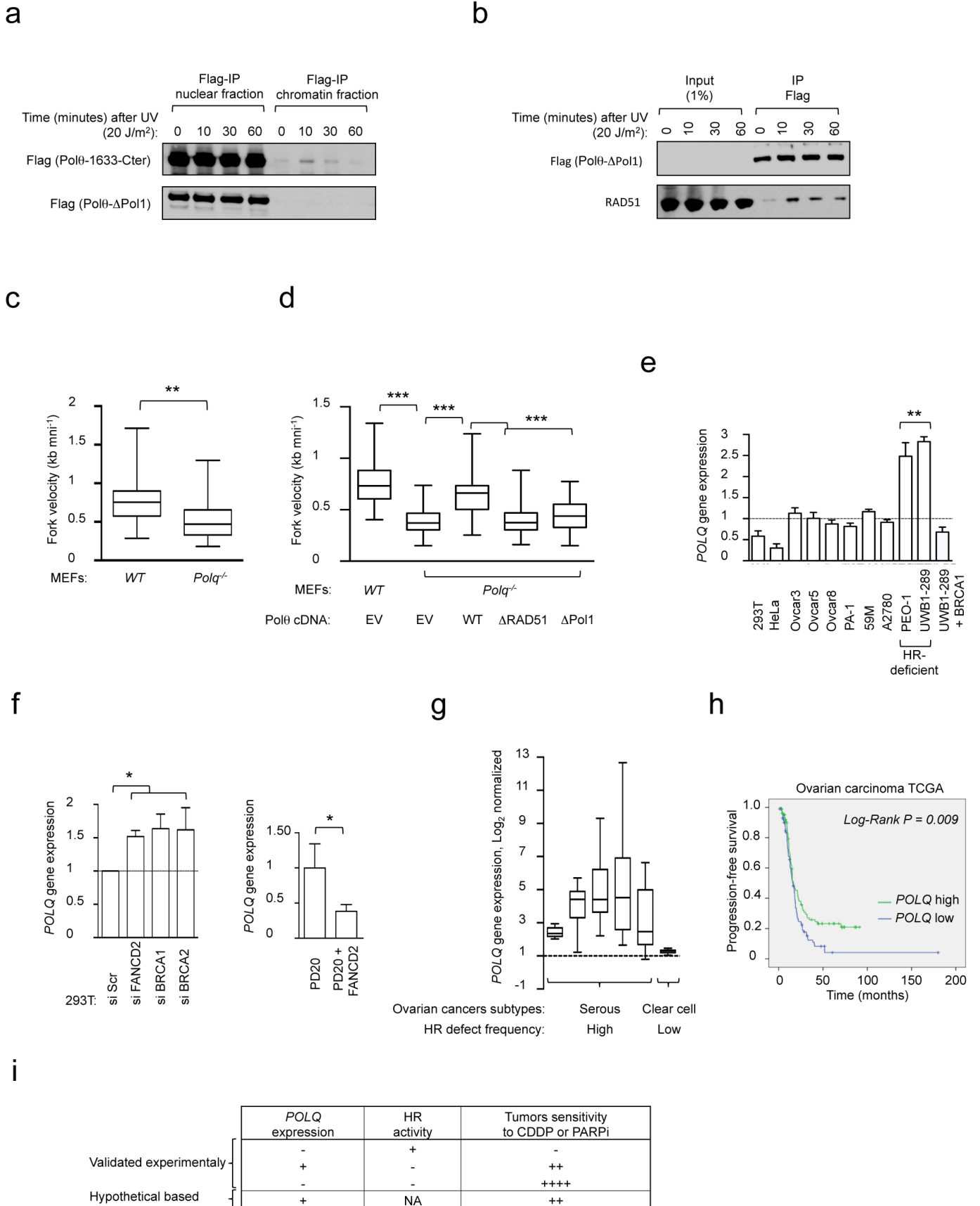


Extended Data Figure 3 | Characterization of RAD51-interacting motifs in Polθ. **a**, Substitution peptide array probed with recombinant RAD51 and analysed by immunoblotting. A 20-mer peptide spanning each of the RAD51 binding sites (shown in Fig. 1g) were created in which each amino acid of the original peptide was mutated to each of the 20 amino acids and RAD51 binding activity was tested. The amino acid change for each of the amino acids of the RAD51 interacting domain of Polθ is shown on the right. Ponceau staining was used to visualize position of the peptides within the array. **b**, GST-RAD51 pull-down with *in vitro* translated Polθ proteins missing

indicated amino acids. **c**, Schematic of Polθ mutants used in complementation studies. **d**, Quantification of IR-induced RAD51 foci in U2OS cells stably integrated with empty vector (EV) or Polθ-ΔPol1 cDNA that is refractory to siPolθ1. Cells were transfected with indicated siRNA and subsequently treated with IR. The number of cells with more than 10 RAD51 foci was calculated relative to control cells (si Scr). **e**, DR-GFP assay in U2OS cells stably integrated with empty vector (EV) or indicated Polθ cDNA constructs refractory to siPolθ1 and transfected with indicated siRNA. All data show mean \pm s.e.m.



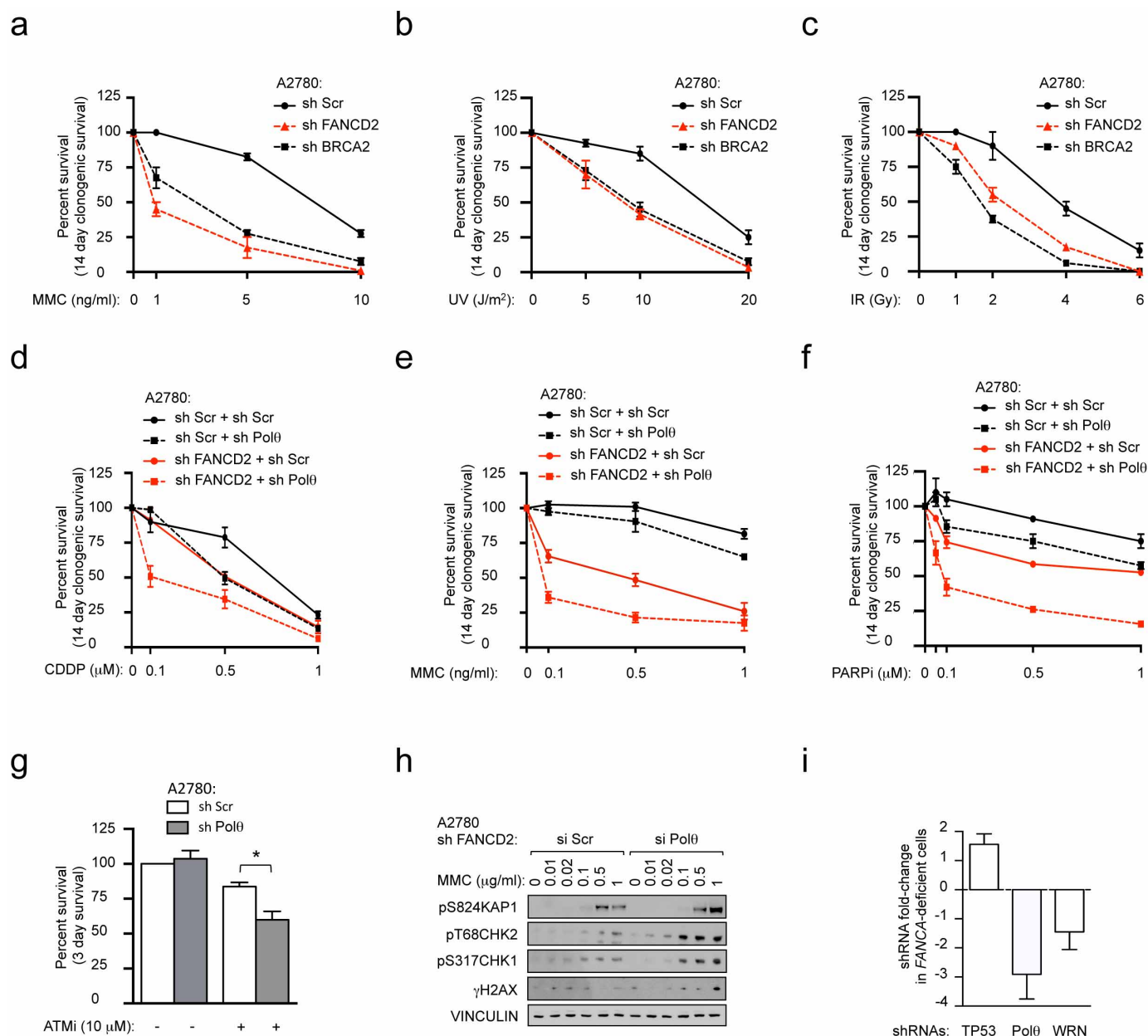
Extended Data Figure 4 | Polθ is an ATPase that suppresses RAD51–ssDNA nucleofilament assembly and formation of RAD51-dependent D-loop structures. **a**, Representative ΔPol2 wild-type radiometric ATPase assay. **b**, Gel mobility shift assays with ΔPol2 wild type and ssDNA. **c**, Coomassie-stained gel showing the purified ΔPol2-A-dead fragment. **d**, Representative ΔPol2-A-dead radiometric ATPase assay. **e**, Quantification of ΔPol2-A-dead ATPase activity. (ssDNA, single-stranded DNA; dsDNA, double-stranded DNA). **f**, Assembly/disruption of RAD51-ssDNA filaments in the presence of increasing amounts of ΔPol2 wild type. The order in which each component was added to the reaction is noted above. **g**, Schematics of the formation of RAD51-dependent D-loop structures. **h**, Formation of RAD51-containing D-loop structures following the addition of increasing amounts of ΔPol2 wild type. **i**, Fraction of D-loop formed following the addition of increasing amounts of ΔPol2 wild type. **j**, Effect of siPolθ and the different Polθ cDNA constructs on HR read-out. NA, not applicable. Data in **i** shows mean ± s.e.m.



Extended Data Figure 5 | Polθ functions under replicative stress and is

induced by HR deficiency. **a**, Polθ recruitment to the chromatin is enhanced by UV treatment. HeLa cells stably integrated with either Flag-tagged ΔPolI or Polθ-1633-Cter (Extended Data Fig. 2g) were subjected to UV treatment. Cells were collected at indicated time points after UV treatment and IPs were performed on nuclear and chromatin fractions. **b**, HeLa cells stably integrated with ΔPolI were treated with UV and collected at the indicated time points following UV exposure. Polθ and RAD51 co-precipitation is enhanced by UV treatment. **c**, Quantification of DNA fibre lengths isolated from wild-type or *Polq*^{-/-} MEFs. **d**, Quantification of DNA fibre lengths isolated from wild-type or *Polq*^{-/-} MEFs transfected with either EV or Polθ cDNA constructs. **e**, *POLQ* gene expression was analysed by RT-qPCR in HR-deficient ovarian cancer cell lines (PEO-1 and UWB1-289) compared with other ovarian cancer cell lines, HeLa (cervical cancer) cells and 293T (transformed human embryonic kidney) cells. Expression was normalized using *GAPDH* gene as a reference. *POLQ* expression values are displayed as fold-change relative to the mean expression in HR-proficient control cells, which was arbitrarily set to 1. **f**, *POLQ* gene expression analysis (RT-qPCR) in 293T cells transfected with siRNA

targeting *FANCD2*, *BRCA1* or *BRCA2* (left panel) and in corrected PD20 cells (PD20 + *FANCD2*) relative to *FANCD2*-deficient cells (PD20) (right panel). Expression was normalized using *GAPDH* gene as a reference. *POLQ* expression values are presented as fold-change relative to the mean expression in control cells, which was arbitrarily set to 1. **g**, *POLQ* gene expression in 5 data sets of serous epithelial ovarian carcinoma (frequently associated with an HR deficiency) and 1 data set of clear cell ovarian carcinoma (subgroup not associated with HR alterations). For each data set, *POLQ* expression values are displayed as fold-change differences relative to the mean expression in control samples, which was arbitrarily set to 1. **h**, Progression-free survival (PFS) after first line platinum chemotherapy for patients with ovarian carcinoma (ovarian carcinoma TCGA). Statistical significance was assessed by the log-rank test ($P < 10^{-2}$). **i**, Effect of Polθ expression levels and HR status on tumour sensitivity to cisplatin or PARPi. NA, not applicable. Box plots in **c**, **d**, and **g** show twenty-fifth to seventy-fifth percentiles, with lines indicating the median, and whiskers indicating the smallest and largest values. Data in **e** and **f** show mean \pm s.e.m.



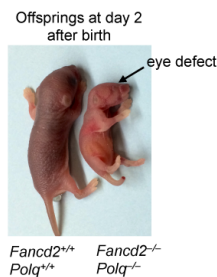
Extended Data Figure 6 | Polθ inhibition sensitizes HR-deficient tumours to cytotoxic drug exposure. **a–c**, Clonogenic formation of A2780 cells expressing scrambled (Scr) shRNA or shRNAs against FANCD2 or BRCA2 with increasing amounts of MMC (**a**), UV (**b**) or IR (**c**). **d–f**, Clonogenic formation of A2780 cells expressing scrambled (Scr) or FANCD2 shRNA, together with shRNA targeting Polθ, in increasing concentrations of CDDP (**d**), MMC (**e**) or PARPi (**f**). **g**, Inhibition of Polθ reduces the survival of A2780 cells after 3 days of continuous exposure to the ATM inhibitor Ku5933. **h**, Immunoblot analyses in A2780 cells expressing FANCD2 shRNA

together with siRNA targeting Polθ or Scr at 24 h after indicated MMC pulse treatment. **i**, FANCA-deficient fibroblasts (GM6418) were infected with a whole-genome shRNA library and treated with MMC for 7 days. The fold-change enrichment of each shRNA after MMC treatment was determined by sequencing relative to the infected cells before treatment. TP53 depletion is known to improve survival of FANCA^{-/-} cells³³. WRN depletion has recently been shown to be synthetically lethal with HR deficiency³⁹. Each column represents the mean of at least 2 independent shRNAs. All data show mean ± s.e.m.

a

Polq status	Fancd2 status	Offspring observed (n)	% observed	% expected	Significant difference
+/+	+/+	19	7.2	6.25	no
+/+	+/-	43	16.2	12.5	no
+/+	-/-	22	8.3	6.25	no
+/-	+/+	36	13.6	12.5	no
+/-	+/-	61	23.0	25	no
+/-	-/-	25	9.4	12.5	no
-/-	+/+	21	7.9	6.25	no
-/-	+/-	34	12.8	12.5	no
-/-	-/-	0 ^u	0	6.25	yes, $P < 0.001$
Total number :		261	100	100	

b



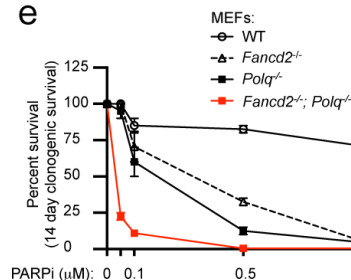
c

Polq status	Fancd2 status	Embryos observed (n)	% observed	% expected	Significant difference
+/+	+/+	13	6.6	6.25	no
+/+	+/-	31	15.7	12.5	no
+/+	-/-	7	3.6	6.25	no
+/-	+/+	22	11.2	12.5	no
+/-	+/-	62	31.5	25	no
+/-	-/-	28	14.2	12.5	no
-/-	+/+	10	5.1	6.25	no
-/-	+/-	16	8.1	12.5	no
-/-	-/-	8	4.1	6.25	no
Total number :		197	100	100	

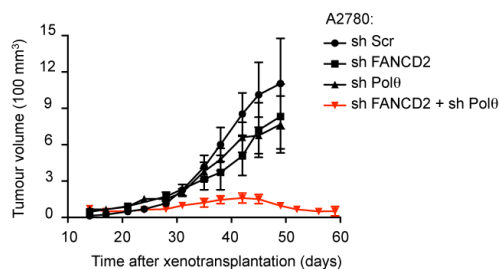
d

Malformation observed	% of <i>Polq</i> ^{-/-} <i>Fancd2</i> ^{-/-} embryos observed with malformations
Reduced body weight	100
Reduced body size	100
Eye defect	100
Limb malformation	12.5

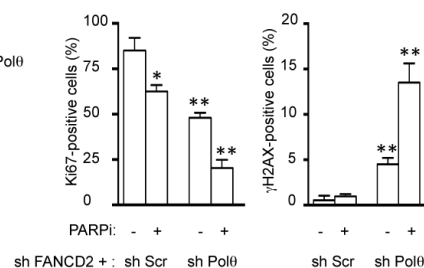
e



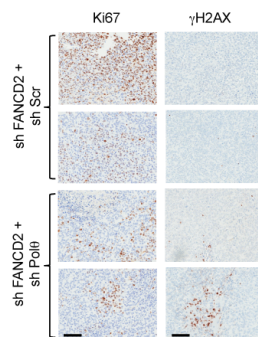
f



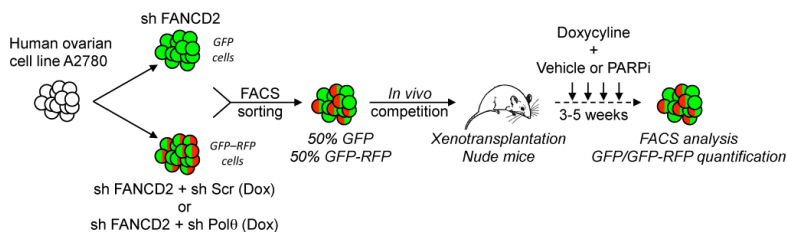
g



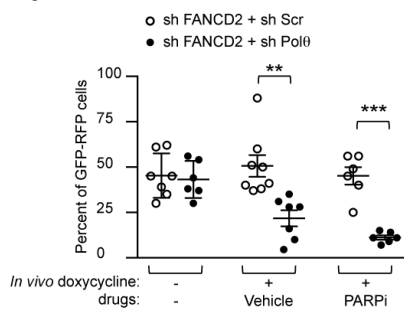
h



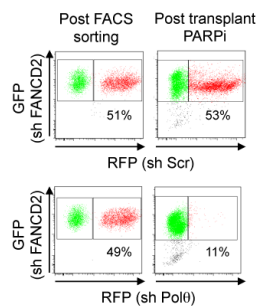
i



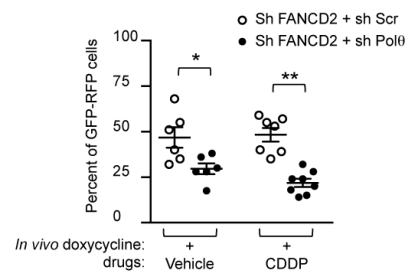
j



k



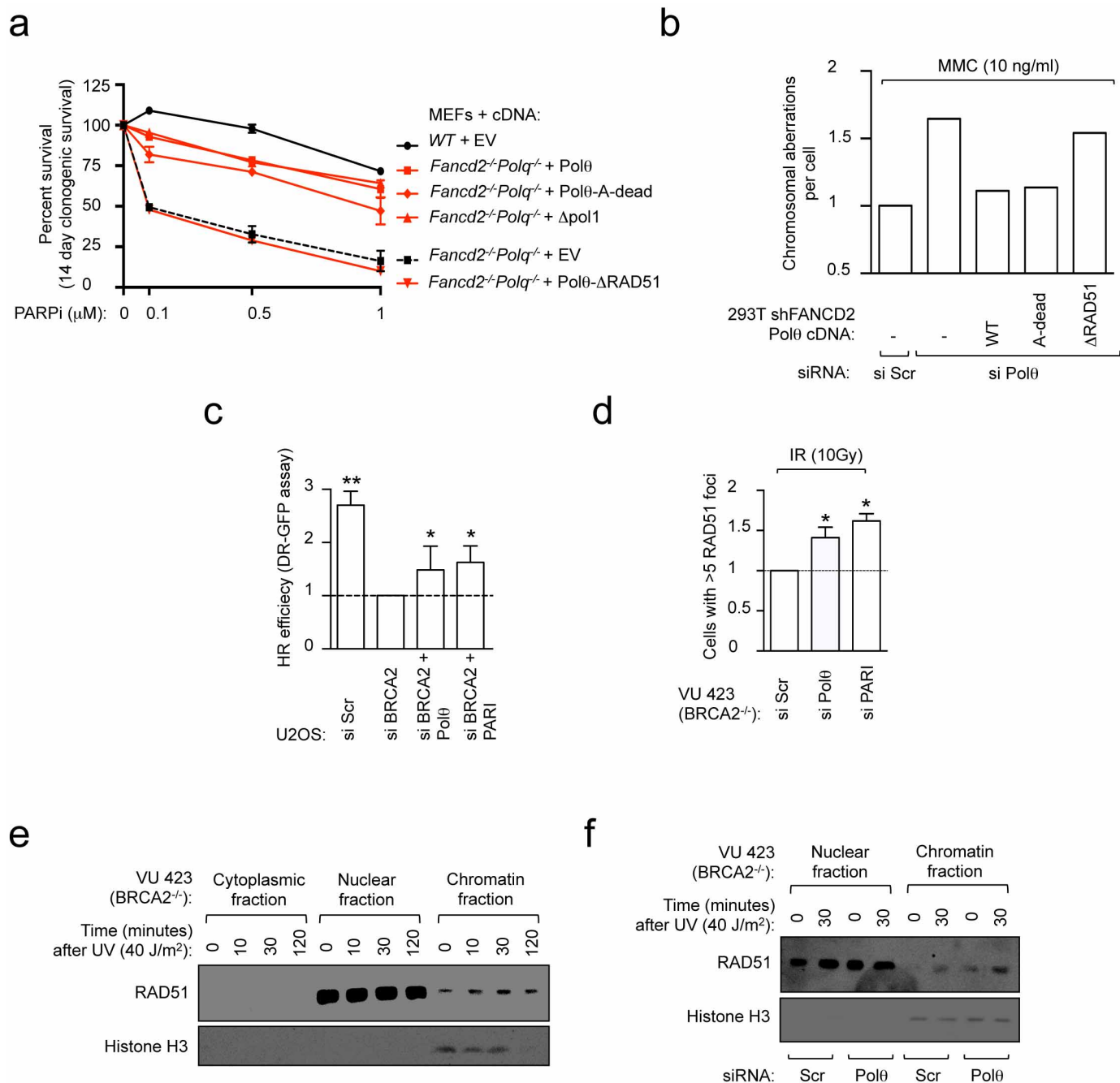
l



Extended Data Figure 7 | HR and Polθ repair pathways are synthetic

lethal *in vivo*. **a**, Genotypes frequencies of offspring from interbred *Fancd2*^{+/-}*Polq*^{+/-} mice. **Ψ**, four *Fancd2*^{-/-}*Polq*^{-/-} offsprings were observed with several congenital malformations and premature death within 48 h of birth. **b**, Description of *Fancd2*^{-/-}*Polq*^{-/-} offspring generated in the study. The offspring presented congenital malformations (that is, eye defects) together with reduced size and body weight. The arrow indicates absence of the right eye. **c**, Genotypes frequencies of E13.5 to E15 embryos (13.5 to 15 days post coitum) from interbred *Fancd2*^{+/-}*Polq*^{+/-} mice. **d**, Description of congenital malformations and their measured frequencies observed in E13.5 to E15 *Fancd2*^{-/-}*Polq*^{-/-} embryos generated in the study. **e**, Clonogenic formation of wild-type, *Fancd2*^{-/-}, *Polq*^{-/-} and *Fancd2*^{-/-}*Polq*^{-/-} MEFs with increasing concentrations of PARPi. **f**, A2780 cells were transduced with indicated shRNAs and xenotransplanted into both flanks of athymic nude mice. The tumour volumes for individual mice were measured biweekly for 8 weeks.

Each group represents $n \geq 5$ tumours from $n \geq 5$ mice. **g**, Ki67 and γ H2AX quantification in tumours treated with either vehicle or PARPi. **h**, Representative Ki67 and γ H2AX staining of A2780-shFANCD2 xenografts expressing sh Scr or sh Polθ in athymic nude mice, treated with either vehicle or PARPi. Scale bars, 100 μ M. **i**, *In vivo* competition assay design. **j**, Tumour chimaerism post-xenotransplantation for indicated conditions. **k**, Representative flow cytometry analysis of tumours before xenotransplantation (post-FACS sorting) or after xenotransplantation (post-transplant, PARPi). The percentage of GFP-RFP positive cells is indicated. **l**, Tumour chimaerism post-xenotransplantation for indicated conditions. For data in **j** and **l**, each circle represents data from one tumour and each group represents $n \geq 7$ tumours from $n \geq 6$ mice. Brackets show mean \pm s.e.m. Data in **e–g** show mean \pm s.e.m. For **f** each group represents $n \geq 6$ tumours from $n \geq 6$ mice.

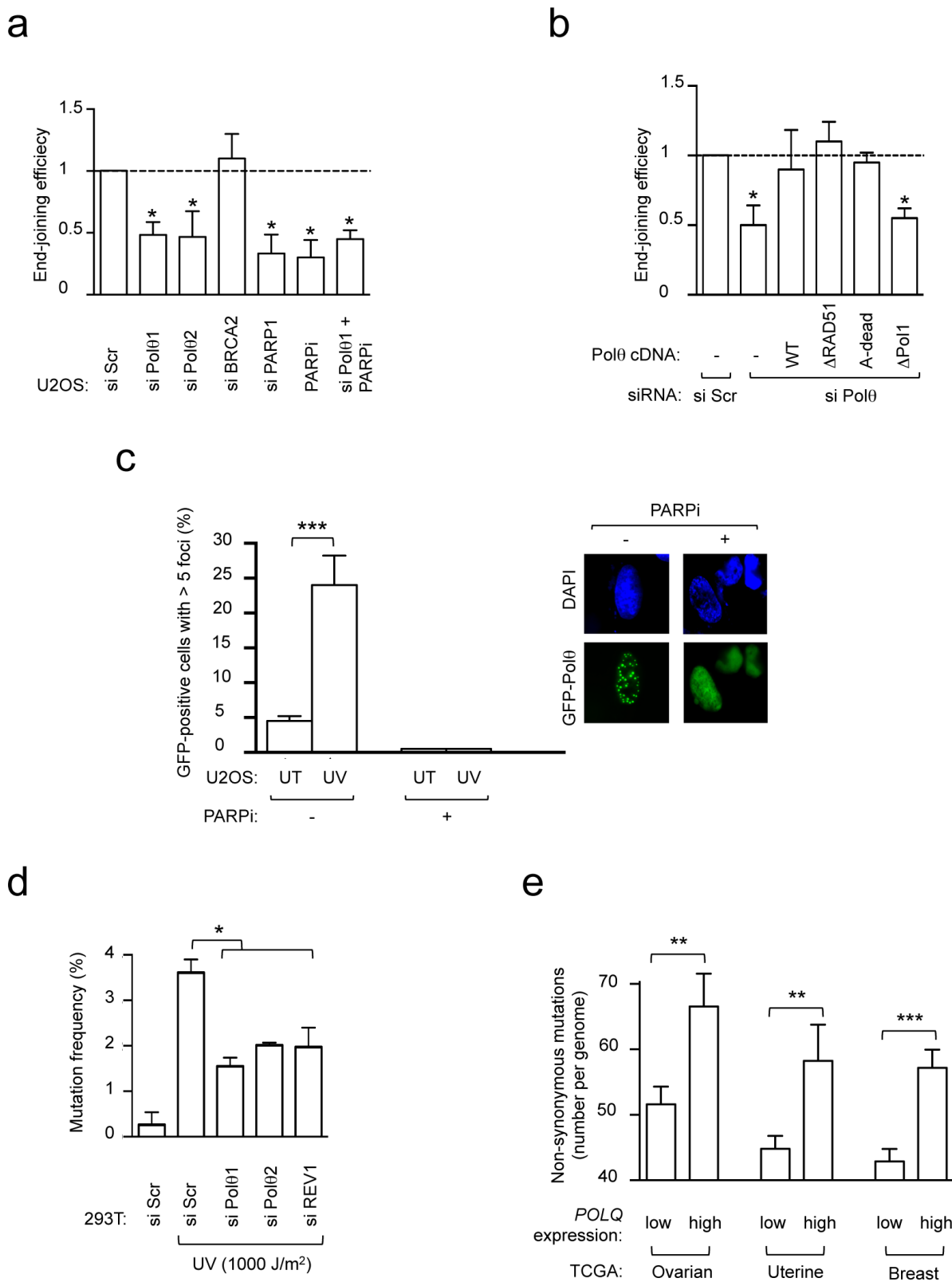


Extended Data Figure 8 | Polθ is required for HR-deficient cell survival and limits the formation of RAD51 structures in HR-deficient cells.

a, Clonogenic formation of *Fancd2*^{-/-}*Polq*^{-/-} MEFs transfected with full-length *POLQ* cDNA constructs in the presence of increasing concentrations of PARPi. **b**, Chromosome breakage analysis of FANCD2-depleted cells that were first transfected with the indicated siRNA and full-length *POLQ* cDNA constructs refractory to siPolθ1 and then exposed to MMC. **c**, DR-GFP assay in U2OS cells transfected with indicated siRNA. **d**, Quantification of baseline and

IR-induced RAD51 foci in U2OS cells transfected with indicated siRNA.

e, RAD51 recruitment to chromatin is enhanced by UV treatment. VU 423 cells (BRCA2^{-/-}) were collected at indicated time points after UV treatment and immunoblotting performed on the cytoplasmic, nuclear and chromatin fractions. **f**, RAD51 recruitment to chromatin in VU 423 cells (BRCA2^{-/-}) transfected with indicated siRNA. Histone H3 was used as a control for chromatin fractionation. All data show mean ± s.e.m.

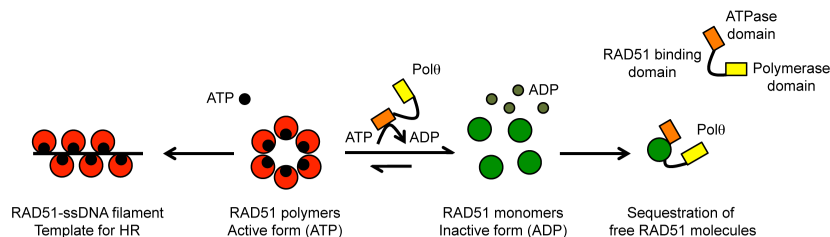


Extended Data Figure 9 | Polθ participates in error-prone DNA repair.

a, End-joining reporter assay in U2OS cells transfected with indicated siRNA and/or treated with PARPi. **b**, End-joining reporter assay in U2OS cells transfected with indicated siRNA and *POLQ* cDNA constructs refractory to siPol01. **c**, UV damage-induced Polθ foci formation in U2OS cells. Polθ foci

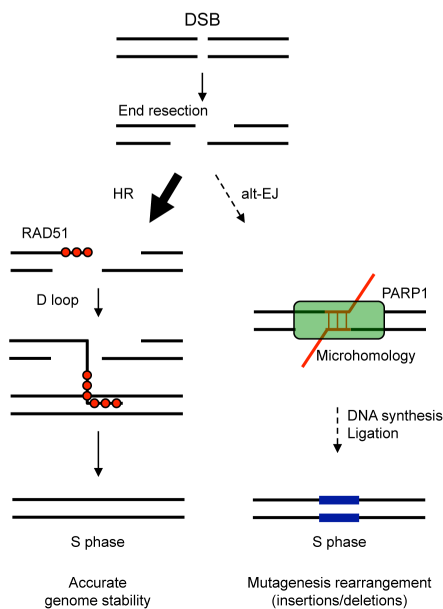
were abolished by pre-treatment with PARPi. UT, untreated. **d**, Mutation frequency was determined in damaged supF plasmid, recovered from siRNA-treated 293T cells. **e**, Non-synonymous mutation count in ovarian, uterine and breast TCGA. All data show mean \pm s.e.m.

a

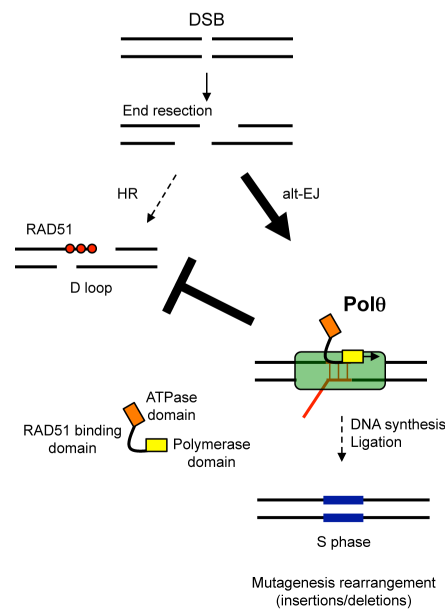


b

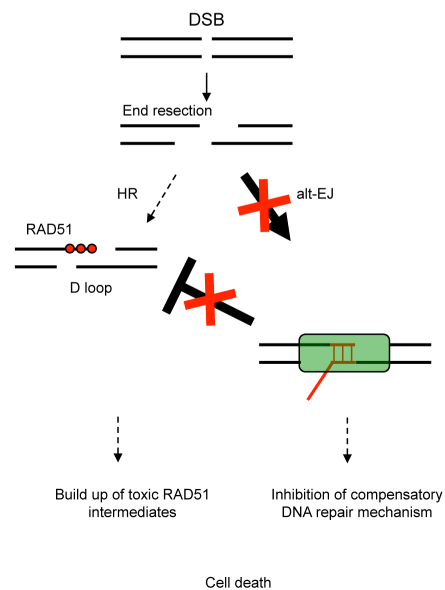
I. Wild type cells



II. HR-deficient cells



III. HR-deficient, Polθ-deficient cells

**Extended Data Figure 10 | Model depicting the role of Polθ in DNA repair.**

a, Mechanistic model for how Polθ limits RAD51-ssDNA filament assembly. According to this model, the ATPase domain of Polθ may prevent the assembly of RAD51 monomers into RAD51 polymers, perhaps by depleting local ATP concentrations. The RAD51 binding domains in the central region of Polθ may then sequester the RAD51 monomers, preventing filament assembly.

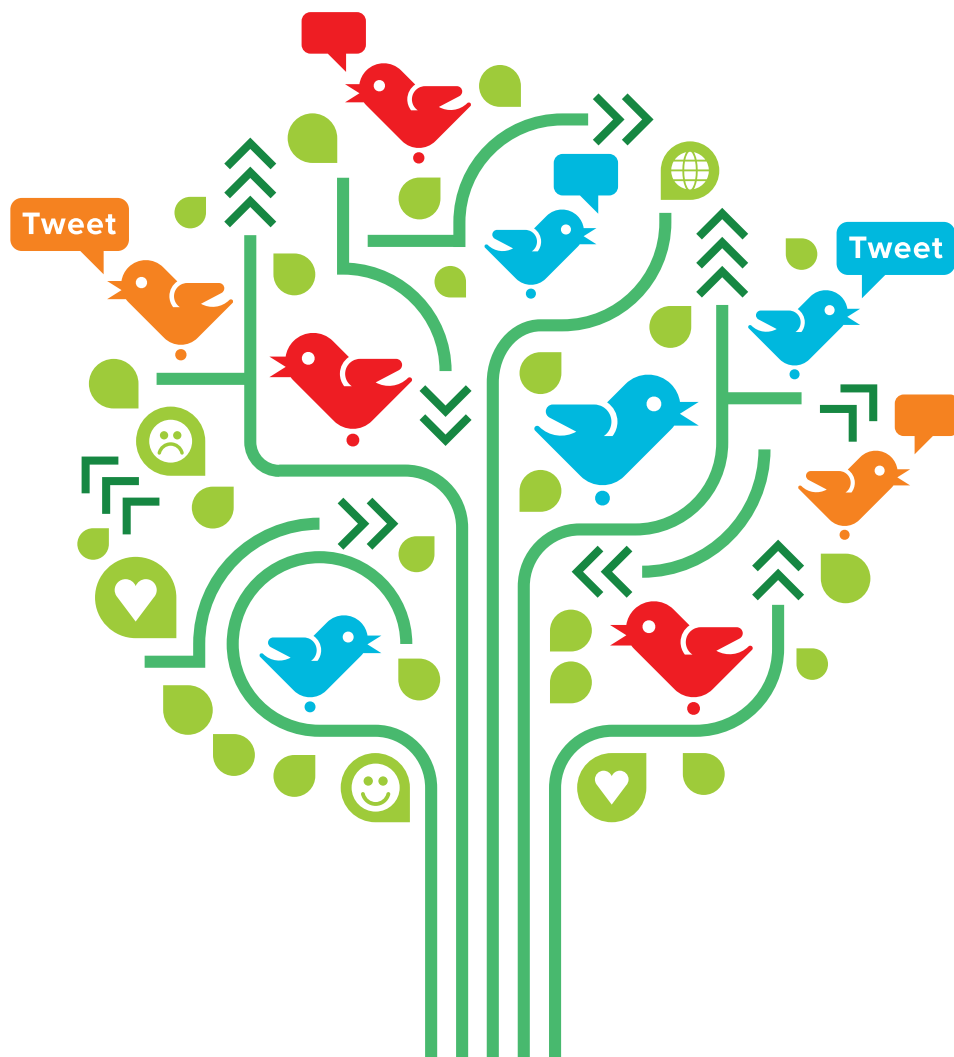
b, (I) Under physiological conditions, Polθ expression is low and its impact on repair of DNA double-strand breaks (DSB) is limited. (II) When HR deficiency occurs, Polθ is then highly expressed and channels DSB repair towards alt-EJ. (III) In the case of an HR-defect, the loss of Polθ leads to cell death through the persistence of toxic RAD51 intermediates and inhibition of alt-EJ.

CAREERS

US FUNDING How state contributions to research and development stack up **p.265**

IMMIGRATION Proposed rules would help scientists to stay in United States **p.265**

NATUREJOBS For the latest career listings and advice www.naturejobs.com



SOCIAL MEDIA

A network boost

How scientists can use Twitter to expand their social contacts and find jobs.

BY MONYA BAKER

Information scientist Cassidy Sugimoto was initially sceptical that Twitter was anything more than a self-promotional time-sink. But when she noticed that her graduate students were receiving conference and co-authoring invitations through connections

made on Twitter, she decided to give the social-media platform a try. An exchange that began last year as short posts, or 'tweets', relating to conference sessions led to a new contact offering to help her negotiate access to an internal data set from a large scientific society. "Because we started the conversation on Twitter, it allowed me to move the conversation into the

physical world," says Sugimoto, who studies how ideas are disseminated among scientists at Indiana University in Bloomington. "It's allowed me to open up new communities for discussions and increase the interdisciplinarity of my research."

Relatively few scientists are taking the opportunities Twitter offers. In a 2014 online *Nature* survey on social media habits, just 12% of the more than 3,000 scientists and engineers who responded reported that they used Twitter regularly (see *Nature* **512**, 126–129; 2014). By contrast, the Pew Research Center, a non-partisan think tank based in Washington DC, found that almost one-quarter of all US adults with Internet access are on Twitter. Researchers in computation-intensive disciplines such as astrophysics tend to use the service more, but no estimates suggest that any discipline of scholars is using Twitter at a higher rate than the general public, says Sugimoto.

That leaves much networking potential untapped, say Twitter enthusiasts. The opportunities for microblogging — posting brief, regular updates online — are plentiful and far-reaching, and can help young scientists to build their careers. Following thought leaders and relevant organizations is an effective, easy way for researchers to learn about important papers, events, funding sources, potential colleagues and job opportunities. Scientists who tweet report that they receive invitations to speak at conferences and events, and make lasting professional connections. There are downsides; scholars need to manage their online time and reputations effectively. Still, by strategically selecting whom to follow and what to contribute on Twitter, young researchers can build a powerful virtual network that will yield opportunities, information and advice.

CREATE CONNECTIONS

People who use Twitter may do so as active participants, posting anything that can fit into 140 characters, and also as followers who read these tweets (see 'On Twitter but not tweeting'). More than one million users follow CERN, the particle-physics laboratory near Geneva, Switzerland, for example. Participants often follow 100 or more users, and so constantly receive posts in their Twitter feed from researchers outside their own immediate networks. Users can also search for posts on a particular topic using the hash symbol followed by a keyword, and curate their connections for discussions relevant to their interests and careers. "By following the people you find interesting and may want to work ►

CHRISTOPHE HEYLEN/GETTY

► with, you're among the first to know when they have an open position within their labs or institutions," says Jacob Jolij, a neuroscientist at the University of Groningen in the Netherlands. For instance, if a potential supervisor or peer is moving lab, he or she may announce it on Twitter. Such posts can suggest that labs and institutions may soon be looking for new employees.

When Thea Whitman was a doctoral student in soil science at Cornell University in Ithaca, New York, in 2013, she forwarded, or retweeted, a post from another soil researcher announcing a tenure-track opening at the University of Wisconsin–Madison. She did not think of herself as a candidate: she was considering options for a postdoctoral position. But when she looked again at the posting, she realized that her interests and qualifications matched the job requirements. She sent in an application, landed an interview and got the job. (She starts in January 2016 after finishing a postdoc stint at the University of California, Berkeley.)

One chemist who writes a blog and tweets under the username Chemjobber about work in the US chemical and drug industry estimates that he tweets three to five positions a day. He receives two or three notes a year from readers who tell him that they found the advert for their new job through one of his posts.

Although plenty of job announcements made on Twitter fail to attract suitable candidates, tweeting and retweeting can help to expand recruitment efforts. In December, Matthew MacManes, a genomic biologist at the University of New Hampshire, Durham, tweeted a link for a tenure-track position in his department. Within 2 weeks, retweets brought the posting to more than 10,000 Twitter users, and some 200 viewed a description of the position, he says. "These are candidates that I wouldn't have otherwise reached."

FIND AN EMPLOYER — OR EMPLOYEE

Twitter is not the primary way that young scientists find jobs, however. Gwynn Benner, who coordinates career services for postdocs and graduate students at the University of California, Davis, says that she often sees a mismatch in Twitter usage at career fairs. "The employers will be tweeting, 'I've got a booth,' and the students are just not on Twitter." A steady stream of tweets come from @naturejobs, @Science-Careers, university career offices, aggregators and employers, but Benner thinks that Twitter could be overwhelming as a primary tool in a job search. Instead, she says, it should be used strategically, to learn what potential employers are up to, and whether they have job openings. "Say there are five companies I want to target," she says. "That's when I get on Twitter."

Twitter can help in early-career researchers' job searches by allowing them to see other users' previous posts and current connections. Arne Bakker at the Stanford University Career Development Center in California provides advice for people with science PhDs and for



Cassidy Sugimoto says that Twitter can help to shape scientific networks.

postdocs, and says that following institutions, companies and individuals on Twitter can offer clues about workplace culture and ongoing projects in a way that static websites do not. That knowledge can be especially helpful during a job interview, he says. It can also paint a picture of what the job might be like, says Whitman. "If your future adviser or colleagues are active on Twitter, it can give you insight into their personality. Do they tend to be negative? Constructive? How do they respond to criticism?"

Employers use Twitter to evaluate potential recruits as well. Evolutionary biologist Iain Couzin is setting up a department to study collective animal behaviour after moving to the Max Planck Institute for Ornithology in Konstanz, Germany. He says that Twitter is becoming a tool to help find excellent young scientists. "I get to know who many of the candidates are as I have also been following them," he says. Danielle Bassett, a bioengineer at the University of Pennsylvania, Philadelphia, says that although she has not used Twitter to recruit lab members directly, she does look at online activity; a history of tweets that demonstrate scientific insight and interdisciplinary interests has increased candidates' chances, she says.

But a Twitter account is not an automatic boost. "Some of my older colleagues think that if you are using social media, you don't have enough to do," says Jessica McCarty, who studies land use at Michigan Tech Research Institute in Ann Arbor. "It is a double-edged sword," warns Jennifer Biddle, an assistant professor at the University of Delaware in Newark who studies environmental microorganisms. "If you are outspoken or mostly post about your personal life, you may create prejudgement."

Or worse: in one particularly controversial case, the University of Illinois at Urbana–Champaign rescinded a job offer for a

tenure-track position after a candidate posted inflammatory tweets. Some scientists, including Chemjobber, have opted to not use their real names to avoid potential conflicts with employers.

S. CRAIG FINLAY

THE NETWORK WAY

Career consultants and university guidance counsellors interviewed by *Nature Careers* de-emphasized Twitter in favour of encouraging an online presence on LinkedIn (see *Nature* **516**, 441–442; 2014). Nonetheless, Twitter and other forms of social media are changing the playing field. Danielle N. Lee, an outreach advocate and postdoc in psychology at Cornell University, says that her blogposts and tweets about making science more inclusive for women and minorities have yielded prestigious speaking engagements and invitations to write articles for publication.

Sugimoto thinks that social media may be starting to reweave the fabric of traditional academic research. "I'm seeing students creating identities that don't have to be routed through the principal investigator," she says. "I see doctoral students making increasing use of Twitter to brand themselves." Although existing studies on the topic are small and research methods are still being worked out, there is some suggestion that social media can have an equalizing effect by making people without access to conventional networks more visible, she says.

Twitter's value to job seekers is more about making connections than finding a newly advertised job, says Chemjobber. "The reason to get on Twitter for your job search is that it offers you a way to short-circuit traditional networking," he says. "It doesn't matter if you're a full professor or a grad student or an early-career person, you can get noticed."

And the social-media platform helps users to cross disciplines, says Hiroki Ueda, a systems biologist at the RIKEN Center for Developmental Biology in Kobe, Japan. "Sometimes I get interested in PhD students

"It doesn't matter if you're a full professor or a grad student or an early-career person, you can get noticed."

and postdocs especially from different fields — chemistry, physics, information science — just through their tweets."

Twitter can enable support networks that would be impossible in the physical world, says Caleph

Wilson, an immunologist at the University of Pennsylvania. He participates in a Twitter group that has launched weekly digital conversations using the hashtag #BlackandSTEM. The platform provides a forum where he can share his experiences as an African American working in science, technology, engineering and mathematics (STEM), and younger scientists can learn from them. "Physically, you may be in a situation where you are the only

#SCIENCETRENDING

On Twitter but not tweeting

A common strategy on Twitter is lurking: reading tweets but not posting them. For many users, Twitter becomes their main way to learn about relevant papers, conferences and news. To build that information feed, users need to choose which streams to follow and what hashtags to monitor, such as #lifeafterPhD. Many follow relevant departments in grant agencies; @NIHfunding has 24,000 followers, for example.

Many journals and journal editors tweet their tables of contents and retweet relevant comments. Beginners on Twitter can also find accounts to follow through retweeted posts and by looking through followers of other users. Lists of recommended people to follow abound as well. Programmes such as TweetDeck or Hootsuite can sort Twitter streams by username and hashtag.

And Twitter is boosting the scope of conferences, too, helping people who cannot attend to follow what is going on. At the Annual Geophysical Union meeting in San Francisco, California, last December, attendees numbered about 24,000, yet more than 28,000 people posted almost 57,000 tweets and retweets with the hashtag #agu14 — double the previous year. Specific sessions within conferences often have their own hashtags, catering to researchers' specific interests.

High levels of Twitter activity can be intimidating, so the best approach is to read tweets selectively. Lisa Balbes, a career-development consultant in Kirkwood, Missouri, advises Twitter users not to even try to check every post. She thinks of Twitter as an additional source of information and networking. "I skim the headlines when I have a couple minutes," she says. **M.B.**

person [in an under-represented group] but through social media, you are in a space where you can have the all-important STEM vent session," he says.

Science exchanges on Twitter are generally convivial, but there is no doubt that Twitter can get ugly. In November, the leader of the Rosetta Mission that landed a probe on a comet wore a shirt printed with scantily clad women. A science writer who tweeted that the attire made astronomy less welcoming to women received multiple tweets telling her to kill herself.

NO PERSONAL POSTS

Although horrible tweets and abusive 'trolls' exist, they are not a significant part of most scientists' experience on Twitter, says Chris Gunter, a researcher and science communicator at Marcus Autism Center and Emory University in Atlanta, Georgia. Those who fear Twitter may not realize how much they can control their experience. "You can unfollow or mute people," she says, "and you can take a break for a while." As a precaution, she avoids inflammatory or overly personal posts, such as using family members' names. For conversations that require nuance, users should switch to other types of communication, she says. It is common for interactions that begin on Twitter to move over to e-mail, for example.

Although conventions on social media are still emerging, the basic rules of networking still apply, says Lisa Balbes, a career-development counsellor in Kirkwood, Missouri. "It's a weird, messy landscape right now," she

says. "It comes down to building a relationship with other people through whatever tools they are using." Relationships require more than a single click. Twitter users should not assume, for example, that being mutual followers with another user means that the person has taken an interest in helping them.

An online reputation for being thoughtful, enterprising and helpful can be as valuable as a long list of publications, says career consultant Peter Fiske, head of PAX Water Technologies in Richmond, California. Scientific conference organizers and observers often follow a meeting's tweetstream to learn what generated excitement, and to find rising stars. Informed tweets can help to draw their attention, says Gunter. For the past several years, she has chaired committees that select speakers and moderators for the American Society of Human Genetics in Bethesda, Maryland. "The tweets alone can't suggest a good speaker," she says, "but tweeting coherently about the topic is always a good sign".

Twitter's greatest advantage may be its flexibility in terms of the time spent and level of commitment. "You can dip your toes in — you don't have to be a crazy twittermaniac," says Titus Brown, a bioinformatician at the University of California, Davis. "In the past few years, I've seen it grow considerably in professional usefulness. It will continue to evolve," he predicts. "Find a way to use it in a way that makes sense with your personality and time constraints — and it will be useful for you." ■

Monya Baker writes and edits for Nature's *Careers* section.

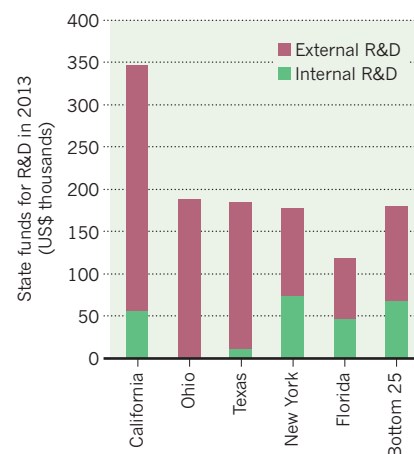
GOVERNMENT FUNDING

State contributions

Individual states funnelled US\$1.8 billion into research labs and studies in 2013, with one-quarter of that devoted to basic research, finds a survey by the National Science Foundation in Arlington, Virginia. Although federal funding for research and development (R&D) dwarfs state investments, state expenditures can help to tailor workforces to regional needs, says James Hearn, associate director of the Institute of Higher Education at the University of Georgia in Athens. Five states together accounted for almost three-fifths of the investments (see 'Top R&D spenders'). External R&D — mainly that at academic institutions — tended to receive more than the internal R&D conducted by state agencies.

TOP R&D SPENDERS

California is top, but even Ohio, Texas and New York spent as much as the bottom 25 states combined.



IMMIGRATION

Scientists gain access

Proposed federal legislation would exempt scientists from some US immigration quotas. Similar legislation introduced by the Senate in 2013 failed to make it through the House of Representatives. However, Atessa Chehraz, an immigration attorney in San Francisco, California, says that foreign researchers would gain many more opportunities to work in the United States if even targeted provisions of the bill pass, such as a proposal to allow graduate students who arrive on non-immigrant visas to seek permanent resident status. Restrictive employment quotas and visas for scientists and other highly trained workers have come under attack in the past decade. More than a dozen higher-education associations are urging Congress to pass the bill.

THE LAST ONE

What a waste.

BY IULIA GEORGESCU

I crash-land amid a cloud of shattered glass. The deserted pedestrian tunnel that runs beneath the streets of Kensington stretches ahead of me. A second later, I am running fast into the walkway's murky depths, thanking science for my light exoskeleton and my employer for paying for it. How did people manage to do their jobs without enhancement prosthetics? The question flits through my mind as I rush past an array of adverts that would have been invisible in the darkness, but for my smart contact lenses. I catch a glimpse of the poster promoting *Extinct! Objects That Are No More*, the new exhibition at the Science Museum. The very place where I have just acquired the item ordered by my employer.

I check my sensors without slowing my pace. No pursuit. Not yet. I tighten my grip on the box in my arms. It had been a risky job. The security at the museum was tight. Very tight. And for what? A gift for his lovely granddaughter, he says. Such a valuable resource wasted as a toy for a five-year-old! Like having a water fight in the desert. I despise my filthy rich employer and his eccentric job requests, but then, he did pay for all of my enhancements. And he got me out of jail. Twice.

I reach the end of the tunnel and stop abruptly. Sensor check. There's some movement at the far end, where I came through the roof. No time to waste. With one hand, I lift a heavy manhole cover and slide into the gap, pulling the lid closed behind me. Hastily, I climb down the ladder, hurrying to the disused tunnels that link to the London Underground system. As I hit the ground, I head for one of the larger passageways and start running again.

Even with all the muscular implants, I am getting tired. No time to stop. It's been a hell of a night. I've been lucky to make it this far. Who would think that a museum would have so many layers of security? Cameras, sensors, drones, guards in the finest exoskeletons. The

irony, of course, is that the objects on display, now worth risking your skin for, were once ubiquitous. The most common things in the world. Even I remember seeing much of that stuff when I was a kid. But after decades of overconsumption and a planet stripped of

are screaming, my vision is getting blurred, but I keep moving.

Part of my mind keeps going back to the exhibition. We made so many wrong choices, wasted so many opportunities, exhausted our resources, lost so much. Just like me and my life. Somehow I never had the strength to do the right things.

I hear the Tube train long before the lights break into the tunnel. A few tens of seconds till it reaches me. I stop and face the oncoming lights. *Focus*. Counting. Three. Two. One. Just before the train hits me, I use my last drops of energy to jump high.

I land on the roof of a carriage. The exoskeleton on my legs cracks badly. I cling to the train with one hand and grip the box with the other. I see the numbers on my retinal display. One minute forty-five seconds to the next stop.

I jump to the platform before the train stops. The badly damaged exoskeleton somehow manages to keep my legs from breaking. Praise the maker. I start running, slower now. I don't even need the sensors to feel my pursuers closing in. I struggle up the stairs and see the faint light of

morning outside. I've made it!

But there's no rescue waiting for me as I emerge into the dawn's rays. Drones close in from behind. Ahead are four armed cops and a little bald man with old-fashioned glasses, whom my retinal display identifies as the museum curator. I spot a different style of drone descending slowly towards me. My bastard employer is abandoning me, but has sent his toy to recover his prize. Fine. Perhaps it's time to quit this job anyway. I gather my little remaining strength to throw the box to the approaching drone.

"Noooo! Don't shoot you idiots!"

I hear the little man screaming and the guns firing. The box shatters and I fall on the cold, wet pavement. Blood tastes warm in my mouth as I hold out my arm. The red helium balloon rises gracefully into the morning sky. ■

Iulia Georgescu is an editor of *Nature Physics*. At work she reads science and on the daily commute on the Tube she imagines the science turning into science fiction.



ILLUSTRATION BY JACEY

resources, stuff that once was worthless is now worth a fortune.

I recheck my sensors. Five pursuers. *Damn*. Perhaps others that are shielded? Should I risk a deeper scan? Better to try to outrun them. If I make it to the meeting point, I might even survive the night. Lucky the box is light.

Run.

Back at the exhibition I think I saw a plastic bottle like my mum used to have. She said that all drinks came in such containers and that cities overflowed with plastic garbage. Hard to imagine that anyone would throw away such valuable material. But that was in the age of waste, before people figured out how badly they needed such resources for the 3D printing industry.

My sensors tell me that the pursuers are closing in. They've brought three drones. I can't outrun those. Desperately, I scan the map on my retinal display. I switch direction and enter another tunnel. Three minutes till the District line train reaches this section. Perhaps I can make it. My muscles